MATHEMATICAL PROBABILITY

UNDERGRADUATE TEXT BOOKS IN MATHEMATICS

By J. G. CHAKRABORTY & P. R. GHOSH

- HIGHER ALGEBRA (Including Modern Algebra)
- ADVANCED HIGHER ALGEBRA
- ANALYTICAL GEOMETRY & VECTOR ANALYSIS
- ADVANCED ANALYTICAL GEOMETRY
- ADVANCED ANALYTICAL DYNAMICS
- DIFFERENTIAL EQUATIONS
- VECTOR ANALYSIS

By B. C. DAS & B. N. MUKHERJEE

- ANALYTICAL DYNAMICS
- DIFFERENTIAL CALCULUS
- INTEGRAL CALCULUS
- HIGHER TRIGONOMETRY
- STATICS
- DYNAMICS

By D. CHATTERJEE

- ELEMENTS OF NUMERICAL ANALYSIS

Published by:

U.N.DHUR & SONS PRIVATE LIMITED

SOME OF OUR REPUTED PUBLICATION BOOKS

REVISED 5TH EDITION IN COMPLIANCE WITH CBCS SYLLABUS OF INDIAN UNIVERSITIES AND EQUIVALENT COURSES, B.TECH, BBA, BCA, & COMPETITIVE EXAMINATIONS

MATHEMATICAL PROBABILITY

A. BANERJEE, M.Sc., D. Phil.

S. K. DE, M.Sc.(*Pure and Applied*), D. Phil.

S. SEN, M.Sc., D. Phil.

U.N.DHUR & SONS PRIVATE LTD.

KOLKATA - 700 073

ISBN 978-93-80673-26-4

First	Published	1999
Second	Edition	2002
	Reprint	2005
Third	Edition	2006
	Reprint	2008
	Reprint	2011
Fourth	Edition	2014
Fifth	Edition	2018
	Reprint	2019
	Reprint	2020
	Reprint	2021

Price : **480/- (Rupees Four Hundred Eighty only)**

Published by

Dr. Purnendu Dhar, M.Sc. (Chem.), Ph.D. *on behalf of*

M/s.U.N.DHUR & SONS PRIVATE LIMITED

2A, Bhawani Dutta Lane, Kolkata - 700 073

Phone : (033) 2241 9573 / 40441734

Mobile : 94330 17104 / 98301 69816

E-mail : **undhur1914@gmail.com**

Website : **www.undhur.com**

Printed by International Printing

60,Hari Ghosh Street, Kolkata - 700 006

Preface To The Fifth Edition

This edition is revised partly with a few additions and alterations to comply with the latest **CBCS** syllabus of all Indian Universities.

Some important problems have been solved by alternative methods, where we felt necessary.

An alternative proof of De Moivre Laplace limit theorem has also been incorporated in the appendix.

Our thanks are due to Dr. Purnendu Dhar, M.Sc. (Chem), Ph.D., who took best care in bringing out this edition in time.

Suggestions for the betterment of the book will be accepted with thanks.

Kolkata
June, 2018

A. Banerjee
S. K. De
S. Sen

Preface To The First Edition

The book entitled "Mathematical Probability and Statistics" is primarily designed to serve as a text book for undergraduate students of Mathematics Honours. It is expected that the book will also be useful to the students of Statistics, Engineering, Commerce and to the candidates for some competitive examinations.

The theory of Probability which provides the foundations of Statistics is presented in the Volume I of the book. This volume contains ten chapters, of which Chapter I deals with some mathematical rudimentaries necessary for the treatment and understanding of the subject in the succeeding chapters.

Chapter II provides an informative introduction concerning the origin and nature of the concept of 'Probability'; the necessity of the axiomatic treatment of the subject has been explained keeping in view the development of the Mathematical Theory of Probability. The remaining eight chapters (Chapter III to X) provide a systematic study of the Mathematical Theory of Probability based on the axiomatic definition of the same.

Best efforts have been made to give the subject a modern touch and in doing so we have confined ourselves to the treatment of the contents of the subject in a precise and rigorous manner. For a clear understanding, detailed explanations have been given whenever felt necessary. At the end of each chapter, except Chapters I and II, sufficient number of problems have been worked out as illustrative examples followed by exercises containing a good number of problems. The problems are selected considering the nature of questions set in the University and other examinations.

Works of eminent authors on the subject have been widely consulted and we express our indebtedness to all of them. In writing Chapter I, standard books on Set Theory, Analysis and Linear Algebra have been availed of.

We are specially indebted to Dr. S. Bakshi of Ramakrishna Mission Vidyamandira, Belur, who went through some chapters of the Manuscript in original and also in its revised form. We have accomodated some of his valuable suggestions and corrections.

Our grateful thanks are also due to Dr. Rabindranath Bhattacharya of Kalyani University, Dr. R. S. Banerjee, Dr. Samir Mukherjee, Dr. Dilip Chakraborty, Dr. Pinaki Ranjan Roy, Dr. Dipti Dutt of Netajinagar Day College, Dr. Y. De, Dr. N. Bhanja of Ramakrishna Mission Residential College, Narendrapur and Dr. S. Mukherjee of Ramakrishna Mission Vidyamandira, Belur for their constant encouragement in completing the work.

Finally, we have to thank Sri Bimalendu Mahalanobis, M. A., for his careful reading of the proof-sheets and rendering many valuable observations.

To conclude, we express our gratitude to the authorities and staff of Messers U. N. DHUR & SONS PRIVATE LIMITED, for their patient and helpful attitude in publishing the work.

Comments and suggestions for the betterment of the book will be gladly accepted. For all its lapses and limitations the authors remain responsible.

Calcutta
January 1999

A. Banerjee
S. K. De
S. Sen

by the Same Author

MATHEMATICAL STATISTICS

CONTENTS

I MATHEMATICAL PRELIMINARIES

II THE CONCEPT OF PROBABILITY

III AN AXIOMATIC CONSTRUCTION OF THE THEORY OF PROBABILITY

IV COMPOUND OR JOINT EXPERIMENT

V PROBABILITY DISTRIBUTION

VI DISTRIBUTION OF MORE THAN ONE DIMENSION

VII MATHEMATICAL EXPECTATION-I

VIII MATHEMATICAL EXPECTATION-II

IX SOME IMPORTANT CONTINUOUS UNIVARIATE DISTRIBUTION

X CONVERGENCE OF A SEQUENCE OF RANDOM VARIABLES & LIMIT THEOREM

CHAPTER I

MATHEMATICAL PRELIMINARIES

1.1. Proposition.

A declarative sentence (not necessarily in the literal sense) which is either true or false but not both is called a statement.

'$2+3=5$' is a statement which is true.

'In Euclidean Geometry the sum of the measures of the angles of a triangle is less than or equal to π' is a statement which is false.

The sentences 'What is the aim of your life ?', 'May God bless you' are not statements'.

We now define the connectives, known as logical connectives, which are used to obtain new statements from a finite number of statements.

Let p, q denote two statements.

(*i*) $p \wedge q$ is a statement which will be read as 'p and q' and which is true if both p, q are true and false otherwise.

(*ii*) $p \vee q$ is a statement which will be read as 'p or q' and which is false if both p, q are false and true otherwise.

(*iii*) $\sim p$ is a statement which will be read as 'not p' and which is true if p is false and false if p is true.

(*iv*) $p \rightarrow q$ is a statement which will be read as 'p implies q' and which is false if q is false when p is true and otherwise $p \rightarrow q$ is true.

(*v*) $p \leftrightarrow q$ is a statement which will be read as 'p implies q and q implies p' and which is true if either p, q are both true or p, q are both false and otherwise $p \leftrightarrow q$ is false.

A statement containing at least one logical connective is called a compound statement, otherwise the statement is called atomic. The statement 'Ram is ill' is atomic whereas the statement '$(2 < x < 3) \wedge$ (x is a rational number)' is a compound statement.

A statement atomic or compound will be called a 'Proposition'.

'$x^2 \geq 0$ for every real number x' is a proposition which is true.

'$\sim$ (2 is an irrational number)' is a proposition which is true.

A proposition concerned with the symbol x is usually denoted as $P(x)$. Here $P(a)$ denotes the proposition obtained from $P(x)$ replacing 'x' by 'a' in every occurrence of x. We can denote the proposition 'for a given real number x, $xy = x$ for every real number y' as $P(x)$. We see that $P(1)$ is false while $P(0)$ is true.

From now on we shall use $p \Rightarrow q$ to mean that the proposition '$p \rightarrow q$' is true and $p \Rightarrow q$ will be read as 'p implies q'. $p \Rightarrow q$ is sometimes read as 'p only if q'.

$p \Leftrightarrow q$ will be used to mean that the proposition '$p \leftrightarrow q$' is true and $p \Leftrightarrow q$ will be read as 'p implies q and q implies p'. $p \Leftrightarrow q$ is also read as 'p if and only if q' or 'q if and only if p'.

1.2. Principle of Mathematical Induction.

Let $P(n)$ be a proposition concerning a natural number n. If $P(1)$ is true and if $P(n) \Rightarrow P(n+1)$ for every natural number n, then $P(n)$ is true for every natural number n.

1.3. Theory of Sets and Related Topics.

Here we shall not attempt to give a precise definition of set by the axiomatic treatment of the theory of sets. Such a definition is not necessary for the development of our subject. Our approach to the primitive notions like collection, elements, membership, etc. related to the concept of a set will be intuitive. Intuitively a set is defined as follows :

"Any well-defined collection of objects of our perception or of our intellect is called a set'.

Here 'collection of objects' is to be regarded as a single entity in the form of a set. By the term 'well-defined' we mean to say that it is possible to speak unambiguously whether an object is a member or not a member of the collection, *i.e.*, there is no doubt about the membership or non-membership. Objects which are members of the collection will be called elements of the set.

Collection of all positive integers is well-defined and so this collection is a set. Collection of all real numbers close to the integer 0 is not well-defined and so this collection is not a set.

The above intuitive notion of set will lead to many paradoxes. Such a paradox will appear if we try to speak of a set which contains everything. This paradox can be avoided by having some set fixed for a given discussion and considering only sets whose members are members of the fixed set.

Sets will be usually denoted by capital letters A, B, C, etc. and members will be denoted by small letters a, b, c, x, y, z, etc. If A be a set and a be an element of the set A, we write $a \in A$ and read 'a belongs to the set A'. If b is not an element of the set A, we write $b \notin A$ and read 'b does not belong to the set A'.

Sets can be described in two ways :

(*i*) *Roster method* :

In this method we describe the set A by listing the elements of the set within { }. If A be the set of all integers lying between 2 and 7, then A can be described as $A = \{3, 4, 5, 6\}$ and in this representation the order in which the elements appear is ignored, so that the same set can also be described as $\{4, 5, 6, 3\}$ or, $\{5, 4, 3, 6\}$ etc.

(*ii*) *Property method* :

Let a set be such that an object is a member of the set if and only if the object has the property P, then the set can be described as $\{x : x$ has the property $P\}$ or as $\{x : P(x)$ is true$\}$, where $P(x)$ denotes the proposition "x has the property P". The set A mentioned above can be described as

$$A = \{x : 2 < x < 7 \text{ and } x \text{ is an integer}\}.$$

Null set :

A set having no element is called a null set or an empty set or a void set and denoted as ϕ or { }.

The set $\{x : x$ is a real root of $x^2 + x + 1 = 0\}$ is a null set.

Subset and Equality : Let A and B be any two sets. A is said to be a subset of B if every element of A is also an element of B

and write $A \subseteq B$ or $B \supseteq A$. So $A \subseteq B$ if for any x, $(x \in A) \Rightarrow (x \in B)$.

If $A \subseteq B$ we also say that A is contained in B or equivalently B contains A.

If $A \subseteq B$ and there exists at least one object b such that $b \in B$ but $b \notin A$, then A is said to be a proper subset of B and we write $A \subset B$ or $B \supset A$.

If $A \subseteq B$ as well as $B \subseteq A$, then the sets A, B are same and in this case we say that A and B are equal and write $A = B$.

Further in this case we say that A is an improper subset of B and also B is an improper subset of A.

From the definition of subset it follows that the null set ϕ is a subset of any set.

For example, the set $\{1, 2, 4\}$ is a proper subset of

$$\{4, -4, 0, 1, 2, 3\}.$$

Some authors write $A \subset B$ whenever A is a subset of B, not necessarily a proper subset of B, and do not use the notation $\subseteq$ or $\supseteq$. Throughout the book we shall use $A \subseteq B$ to mean that A is a subset of B.

Union and Intersection :

Let A and B be any two sets. The union (or join) of A and B, denoted by $A \cup B$, is a set defined by

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Using logical connectives, $A \cup B$ can be described as

$$\{x : (x \in A) \vee (x \in B)\}.$$

The intersection (or meet) of A and B, denoted as $A \cap B$, is a set defined by

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Using connectives, $A \cap B$ can be described as

$$\{x : (x \in A) \wedge (x \in B)\}.$$

If $A \cap B = \phi$, then A, B are said to be disjoint.

For example, if $A = \{1, 0, -1, 2\}$, $B = \{0, 1, 4\}$, then

$$A \cup B = \{1, 0, -1, 2, 4\} \text{ and } A \cap B = \{1, 0\}.$$

Universal set : At the beginning we mentioned that for avoiding paradoxes, in any given discussion with sets we keep some set fixed such that we can assume that any set under discussion (unless otherwise stated) would be a subset of the fixed set. This fixed set will be called universal set.

If set theory be applied in studying the algebra of real numbers then the set of all real numbers is taken as universal set, whereas in studying plane Euclidean Geometry the set of all points in Euclidean plane is taken as universal set. The universal set is usually denoted by X or S or U.

Difference of two sets :

Let A and B be two sets. The difference of the sets A and B, denoted as $A - B$, is defined by

$$A - B = \{a : (a \in A) \wedge (a \notin B)\}.$$

For example, let $A = \{1, 2, -1, 4\}$ and $B = \{1, 4, 0, 3, 8\}$.

Then $A - B = \{2, -1\}$, $B - A = \{0, 3, 8\}$.

Complement of a set :

Let S be the universal set and A be any set. The set $S - A$ is called the complement of the set A and it is usually denoted by A^c or $\bar{A}$ or A'. We shall denote the complement of A by $\bar{A}$.

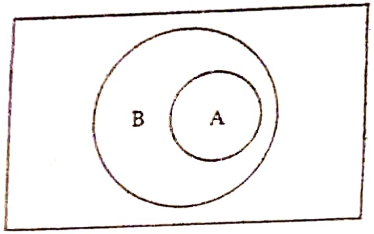
Thus $\bar{A} = \{a : (a \in S) \wedge (a \notin A)\}$.

Venn Diagram :

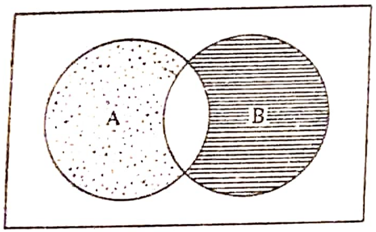
Sets and subsets and their union, intersection, difference, complement may all be represented by diagrams known as Venn diagrams after the name of British logician John Venn (1834—1923).

The universal set X is represented by the region enclosed by a rectangle and a set is represented by its subregion enclosed by a closed curve, usually a circle which lies entirely within the region representing X (with an exception for representing the complement of a set). Subset of a set A is represented by a subregion enclosed by a closed curve lying entirely within the region representing A. The intersection of two subsets is denoted by the region common to the two regions which represent the two sets. Likewise, the union, difference, complements are obtained as different regions

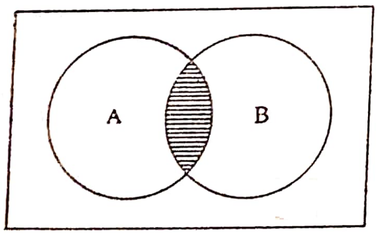
within the rectangle representing X. These are illustrated in the following diagrams :



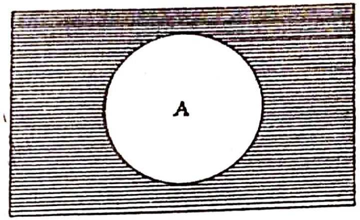
A is a subset of B.
Fig. 1.3.1



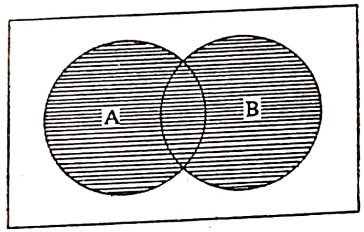
The dotted region denotes $A—B$ and the region shaded by horizontal lines represent $B—A$.
Fig. 1.3.2



Shaded region represents $A \cap B$.
Fig. 1.3.3



Shaded region represents complement $\bar{A}$ (Here we note that the set $\bar{A}$ is not represented by subregion of X enclosed by a closed curve as mentiond earlier).
Fig. 1.3.4



Shaded region represents $A \cup B$.
Fig. 1.3.5

Cartesian Product of two non empty Sets :

Let A and B be two non empty sets. The Cartesian product of A and B, denoted as $A \times B$, is defined by

$$A \times B = \{(a, b) : (a \in A) \wedge (a \in B)\}.$$

where the object (a, b) is called an ordered pair which means intuitively a collection of the objects a, b with an idea of order that a is the first object and b is the second.

For example let $A = \{1, 2, 3\}$, $B = \{0, -1\}$, then

$$A \times B = \{(1, 0), (1, -1), (2, 0), (2, -1), (3, 0), (3, -1)\};$$

$$B \times A = \{(0, 1), (-1, 1), (0, 2), (-1, 2), (0, 3), (-1, 3)\}.$$

Here $A \times B$ and $B \times A$ are disjoint.

Power set of a set : Let A be a given set. The power set of the set A, denoted as $P(A)$, is a set defined by

$$P(A)=\{B : B \subseteq A\}, \text{ i.e., } P(A) \text{ is the set of all sub-sets of } A.$$

For example, if $A=\{0, 1\}$, then $P(A)=\{\{0\}, \{1\}, \phi, \{0, 1\}\}$, which has 4 distinct elements.

Fundamental laws regarding Union, Intersection and Complement of sets :

1. *Commutative laws* :
   (a) $A \cup B=B \cup A$
   (b) $A \cap B=B \cap A$

where A and B are any two sets.

2. *Associative laws* :
   (a) $A \cup (B \cup C)=(A \cup B) \cup C$
   (b) $A \cap (B \cap C)=(A \cap B) \cap C$

for any three sets A, B, C.

3. *Distributive laws* :
   (a) $A \cap (B \cup C)=(A \cap B) \cup (A \cap C)$
   (b) $A \cup (B \cap C)=(A \cup B) \cap (A \cup C)$

for any three sets A, B and C.

4. *Absorption laws* :
   (a) $A \cup (A \cap B)=A$
   (b) $A \cap (A \cup B)=A$

for any two sets A and B.

5. *Idempotent laws* :
   (a) $A \cup A=A$
   (b) $A \cap A=A$

for any set A.

6. *Laws of complement* :
   (a) $A \cup \bar{A}=X$
   (b) $A \cap \bar{A}=\phi$

for any set A.

7. *Laws of identity* :
   (a) $A \cup \phi=A$
   (b) $A \cap X=A$

for any set A.

8. *De Morgan's laws* :
   (a) $\overline{(A \cup B)}=\bar{A} \cap \bar{B}$
   (b) $\overline{(A \cap B)}=\bar{A} \cup \bar{B}$

for any two sets A and B.

Another important property regarding complements and sub-sets :

9 (a) If $A \subseteq B$, then $\bar{A} \supseteq \bar{B}$.
9 (b) If $\bar{A} \supseteq \bar{B}$, then $A \subseteq B$.

Let us prove 8(a).

For arbitrary x, $x \in \overline{(A \cup B)} \Rightarrow x \notin A \cup B$ and $x \in X$

$\Rightarrow (x \notin A) \wedge (x \notin B)$ and $x \in X$

(for, if x belongs to any one of the sets A, B, then x will belong to $A \cup B$)

$\Rightarrow (x \in \bar{A}) \wedge (x \in \bar{B})$

$\Rightarrow x \in \bar{A} \cap \bar{B}$

Thus, $x \in \overline{(A \cup B)} \Rightarrow x \in (\bar{A} \cap \bar{B})$. $\therefore$ $\overline{(A \cup B)} \subseteq (\bar{A} \cap \bar{B})$.

Again for arbitrary y, $y \in \bar{A} \cap \bar{B} \Rightarrow (y \notin A) \wedge (y \notin B)$ and $y \in X$

$\Rightarrow y \notin (A \cup B)$ and $y \in X \Rightarrow y \in \overline{(A \cup B)}$.

Thus, $y \in \bar{A} \cap \bar{B} \Rightarrow y \in \overline{(A \cup B)}$. $\therefore$ $(\bar{A} \cap \bar{B}) \subseteq \overline{(A \cup B)}$.

Hence we have shown that

$\overline{(A \cup B)} \subseteq (\bar{A} \cap \bar{B})$ and $(\bar{A} \cap \bar{B}) \subseteq \overline{(A \cup B)}$.

Therefore, $\overline{(A \cup B)}=(\bar{A} \cap \bar{B})$.

Similar proofs can be given for the other fundamental laws.

We can also verify the law 8(a) by Venn Diagram.

First shade $A \cup B$ with horizontal lines. Then $\overline{(A \cup B)}$ is the region outside $A \cup B$, shaded by upward slanted lines (Fig. 1.3.6)

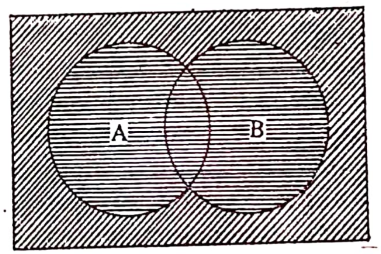
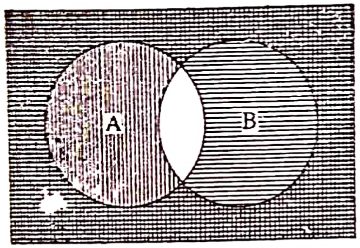


Fig. 1.3.6 $\overline{A \cup B}$



Next we shade $\bar{A}$ and $\bar{B}$ shaded respectively by horizental and vertical lines. Then cross-hatched area is $\bar{A} \cap \bar{B}$. (Fig. 1.3.7)

Notice that $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$.

Fig. 1.3.7 $\bar{A} \cap \bar{B}$

Principle of Duality:

The fundamental laws 1—8 regarding union, intersection, complement have been listed in pairs, each law in a pair can be deduced from the other by interchanging $\cup$ and $\cap$, ϕ and X, and for the pair of laws 9 (a), 9 (b) each can be obtained from the other by interchanging $\subseteq$ and $\supseteq$, A and $\bar{A}$, B and $\bar{B}$.

'If P be a proposition which is true and $P°$ be the proposition obtained from P by interchanging $\cup$ and $\cap$, ϕ and X, $\subseteq$ and $\supseteq$, A and $\bar{A}$, B and $\bar{B}$, as mentioned above then $P°$ is also true.' This is known as principle of duality.

For example, interchanging $\cup$ and $\cap$ we get 8(b) from 8(a) and conversely.

Mapping or Function:

Let A, B be two non-empty sets. A mapping f of A to B, denoted as $f: A \to B$ gives a correspondence between the elements of A and B such that for every element $a \in A$, there exists a unique element $b \in B$ under the correspondence. So formally mapping $f: A \to B$ can be defined as follows:

A subset f of $A \times B$ is called a mapping of A to B if for every element $a \in A$, there exists a unique element $b \in B$ such that the ordered pair $(a, b) \in f$.

If, for the mapping $f: A \to B$, $(a, b) \in f$, we write $f(a) = b$ and say that, b is the image of a under f. Here the set A is called the domain and the set B is called the co-domain of the mapping $f: A \to B$.

The set $\{f(a) : a \in A\}$ is a subset of B and it is called the range of the mapping $f: A \to B$ and it is denoted by $f(A)$.

If $f(A) = B$, then the mapping $f: A \to B$ is called a surjective or onto mapping.

A mapping $f: A \to B$ is called injective or one-one if $[f(a_1) = f(a_2)] \Rightarrow (a_1 = a_2)$, or equivalently, $(a_1 \neq a_2) \Rightarrow [f(a_1) \neq f(a_2)]$, where $a_1, a_2 \in A$.

A mapping $f: A \to B$ is called bijective if it is both injective and surjective.

Inverse of a mapping:

Let $f: A \to B$ be a mapping. If there exists a mapping $g: B \to A$ such that $g(b) = a$ if and only if $f(a) = b$, where $a \in A$, $b \in B$, then $g: B \to A$ is called the inverse of $f: A \to B$.

It can be shown that the inverse of a mapping $f: A \to B$ exists and is unique if and only if f is bijective. Further, the inverse mapping (if it exists) is also bijective. The inverse of $f: A \to B$ is denoted by $f^{-1}: B \to A$.

If $f: A \to B$ be bijective, we say that A is in bijective correspondence with B.

As an example of mapping, let $A = \{1, 0, -1\}$, $B = \{2, 1, 0\}$. A mapping $f: A \to B$ is defined by $f(a) = a + 1$, where $a \in A$. Here $f(1) = 2$, $f(0) = 1$, $f(-1) = 0$. So here $f(A) = B$ and hence f is surjective. Also $f(0)$, $f(1)$, $f(-1)$ are all distinct and consequently f is injective. So $f: A \to B$ is bijective.

The inverse $f^{-1}: B \to A$ of the above mapping exists and it is defined by $f^{-1}(b) = b - 1$, where $b \in B$. Here $f^{-1}(1) = 0$, $f^{-1}(2) = 1$ and $f^{-1}(0) = -1$.

Finite set and Infinite set:

Intuitively we can say that a set is finite if it has finite number of distinct elements. Precisely a finite set is defined as follows:

A non-empty set A is called a finite set if there exists a positive integer n such that there is a bijective mapping of the set $J(n)$ onto A, where $J(n) = \{1, 2, 3, \ldots\ldots, n\}$. Here the positive integer n is called the number of distinct elements of the set A. The null set ϕ is also taken as a finite set. The number of elements of the

null set is zero. Now if A be any finite set, we shall denote the number of distinct elements of A by $n(A)$. It can be shown easily by Venn Diagram that

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

for any two sets A and B.

A set is called an infinite set if it is not a finite set. It can be shown that a set B is infinite if and only if it is in bijective correspondence with a proper subset of itself.

For example, let N be the set of all natural numbers and let $A = N - \{1\}$. The mapping $f: N \rightarrow A$, where $f(n) = n+1$, $n \in N$, is a bijective mapping. Here A is a proper subset of N. So N is an infinite set.

Countable set :

A set A is called a countable set or an enumerable set if there exists a bijective mapping of N onto A, where N is the set of all natural numbers.

It can be shown that the set of all integers and the set of all rational numbers are examples of countable sets.

An infinite set A is called uncountable or non-enumerable if A is not countable.

It can be shown that the set of all real numbers is uncountable.

A set is called at most countable if it is either finite or countable.

Union and intersection of arbitrary collection of sets :

Let σ be an arbitrary set called an index set (index set is not necessarily a subset of the universal set X). Then $\tau = \{A_\alpha : \alpha \in \sigma\}$ is a collection of sets, where each $A_\alpha \subseteq X$. The union of the members of τ, denoted as $\bigcup_{\alpha \in \sigma} A_\alpha$, is defined as

$$\bigcup_{\alpha \in \sigma} A_\alpha = \{x : x \in A_\alpha \text{ for at least one } \alpha \in \sigma\}.$$

The intersection of the members of τ, denoted as $\bigcap_{\alpha \in \sigma} A_\alpha$, is defined as

$$\bigcap_{\alpha \in \sigma} A_\alpha = \{x : x \in A_\alpha \text{ for every } \alpha \in \sigma\}.$$

We observe that these definitions are consistent with the definitions of union and intersection of two sets.

If σ be a null set, we agree to define $\bigcup_{\alpha \in \sigma} A_\alpha$ as the null set and $\bigcap_{\alpha \in \sigma} A_\alpha$ as the universal set X.

If σ be a finite set containing n distinct elements, say $1, 2, \ldots\ldots, n$, then the union and intersection of the sets $A_1, A_2, \ldots\ldots, A_n$ are respectively denoted as $\bigcup_{\alpha=1}^{n} A_\alpha$ and $\bigcap_{\alpha=1}^{n} A_\alpha$.

If σ be a countable set, then we say that τ is a collection of countably infinite number of sets and in this case the union and intersection are respectively denoted as $\bigcup_{\alpha=1}^{\infty} A_\alpha$ and $\bigcap_{\alpha=1}^{\infty} A_\alpha$.

From now we shall denote the null set ϕ by O, the union of A and B by $A+B$, the intersection of A and B by AB. The union and intersection of finite number of sets $A_1, A_2, \ldots\ldots, A_n$ will then be denoted as $\sum_{\alpha=1}^{n} A_\alpha$ and $\prod_{\alpha=1}^{n} A_\alpha$ respectively and with these notations, the union and intersection of countably infinite number of sets are respectively denoted as $\sum_{\alpha=1}^{\infty} A_\alpha$ and $\prod_{\alpha=1}^{\infty} A_\alpha$.

To conclude, we mention here that De Morgan's laws hold for any arbitrary collection of sets. In particular, for countably infinite number of sets,

$$\overline{\left(\sum_{n=1}^{\infty} A_n\right)} = \prod_{n=1}^{\infty} \bar{A}_n \text{ and } \overline{\left(\prod_{n=1}^{\infty} A_n\right)} = \sum_{n=1}^{\infty} \bar{A}_n.$$

Functions of a single real variable :

The functions studied in elementary calculus are particular mappings, where the domain A and co-domain B are subsets of R, R being the set of all real numbers. Here the function is called a real valued function of a real variable and it is usually expressed as $y = f(x)$ instead of $f: A \rightarrow B$.

We assume that that the readers are well acquainted with the concepts of limit, continuity, differentiability, etc. of real valued

functions of a single real variable. In particular, we mention the following theorem in connection with the existence of inverse function of a real valued function of a single real variable :

'Let $f(x)$ be a real valued function of a single real variable x with its domain $E \subseteq R$ and range $X \subseteq R$, where R is the set of all real numbers. If $f(x)$ is strictly monotonic increasing (decreasing) on E, then the inverse function f^{-1} exists which is also monotonic increasing (decreasing) on its domain X'.

Some notations with reference to the set of all real numbers :

From now we shall denote the set of all real numbers by R. Let a, b be two real numbers where $a < b$. The set $\{x : a \leq x \leq b\}$ is denoted as $[a, b]$ and is called a closed interval. The set $\{x : a < x < b\}$ is called an open interval and is denoted by (a, b). The set $\{x : a < x \leq b\}$ is denoted by $(a, b]$ and the set $\{x : a \leq x < b\}$ by $[a, b)$ and are called half open intervals. The set R can also be denoted by $(-\infty, \infty)$.

1.4. Functions of Several Real Variables.

Let $u = f(x, y)$ represent a real valued function of two variables defined in a domain D, where $D \subseteq R \times R$. Let $(a, b) \in D$. The function $f(x, y)$ is said to be continuous at (a, b) if corresponding to every positive real number ϵ, there exists a positive real number δ such that $|f(x, y) - f(a, b)| < \epsilon$ whenever $|x - a| < \delta$, $|y - b| < \delta$ and $(x, y) \in D$.

The function f is said to be continuous on D if it is continuous at every point of D. We can similarly define continuity of a function of three or more variables.

Partial Derivatives :

For the function $u = f(x, y)$, the partial derivatives $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ are defined as

$$\frac{\partial u}{\partial x} = \underset{h \to 0}{Lt} \frac{f(x+h, y) - f(x, y)}{h}$$

$$\frac{\partial u}{\partial y} = \underset{k \to 0}{Lt} \frac{f(x, y+k) - f(x, y)}{k}$$

provided the limits exist finitely. Higher order derivatives $\frac{\partial^2 u}{\partial x^2}$, $\frac{\partial^2 u}{\partial y^2}$, $\frac{\partial^2 u}{\partial x \partial y}$, $\frac{\partial^2 u}{\partial y \partial x}$, $\frac{\partial^3 u}{\partial x^3}$ etc. are defined by

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right), \frac{\partial^2 u}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right), \frac{\partial^2 u}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right) \text{ etc.}$$

A sufficient condition for the equality of $\frac{\partial^2 u}{\partial x \partial y}$ and $\frac{\partial^2 u}{\partial y \partial x}$: If the function defined by $u = f(x, y)$ be such that (*i*) $\frac{\partial^2 u}{\partial x \partial y}$ is continuous at (a, b), (*ii*) $\frac{\partial u}{\partial x}$ exists in a neighbourhood of (a, b), then $\frac{\partial^2 u}{\partial y \partial x}$ exists at (a, b) and $\frac{\partial^2 u}{\partial x \partial y} = \frac{\partial^2 u}{\partial y \partial x}$ at (a, b).

[By a neighbourhood of (a, b) we mean a set of the type : $\{(x, y) : a - \delta < x < a + \delta, b - \delta < y < b + \delta$, where $\delta > 0\}$]

Jacobian :

Let $u_i = f_i(x_1, x_2, \ldots\ldots, x_n)$, for $i = 1, 2, \ldots\ldots n$, represent n real valued functions of n real variables $x_1, x_2, \ldots\ldots, x_n$. Then the Jacobian of $u_1, u_2, \ldots\ldots, u_n$ with respect to $x_1, x_2, \ldots\ldots, x_n$, denoted by $\frac{\partial(u_1, u_2, \ldots\ldots, u_n)}{\partial(x_1, x_2, \ldots\ldots, x_n)}$, is defined as the value of the determinant

$$\begin{vmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \cdots\cdots & \frac{\partial u_1}{\partial x_n} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \cdots\cdots & \frac{\partial u_2}{\partial x_n} \\ \cdots & \cdots & & \cdots \\ \cdots & \cdots & & \cdots \\ \frac{\partial u_n}{\partial x_1} & \frac{\partial u_n}{\partial x_2} & \cdots\cdots & \frac{\partial u_n}{\partial x_n} \end{vmatrix}$$

, provided the partial derivatives exist.

It can be shown that

$$\frac{\partial(u_1, u_2, \ldots\ldots u_n)}{\partial(x_1, x_2, \ldots\ldots, x_n)} = \frac{1}{\frac{\partial(x_1, x_2, \ldots\ldots, x_n)}{\partial(u_1, u_2, \ldots\ldots, u_n)}}$$

provided $\frac{\partial(x_1, x_2, \ldots\ldots, x_n)}{\partial(u_1, u_2, \ldots\ldots, u_n)} \neq 0$.

Criterion for existence of extreme values of a function of three variables :

Let $f(x, y, z)$ represent a function where $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$ vanish at (a, b, c). The function $f(x, y, z)$ has a maximum at (a, b, c) if

(*i*) $\frac{\partial^2 f}{\partial x^2} < 0$

(*ii*) $\begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{vmatrix} > 0$

(*iii*) $\begin{vmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{vmatrix} < 0$

at (a, b, c). On the other hand, $f(x, y, z)$ has a minimum at (a, b, c) if the values of the expressions (*i*), (*ii*). (*iii*) are all positive at (a, b, c).

1.5. Step Function.

Let $f : [a, b] \to R$ be a function. The real valued function f is said to be a step function defined on $[a, b]$, if there exist finite number of points $c_1, c_2, \ldots\ldots, c_n \in (a, b)$ such that

$$\begin{aligned} f(x) &= k_0 \text{ if } a \leq x < c_1 \\ &= k_0 + k_1 \text{ if } c_1 \leq x < c_2 \\ &= k_0 + k_1 + k_2 \text{ if } c_2 \leq x < c_3 \\ &\ldots \quad \ldots \quad \ldots \\ &\ldots \quad \ldots \quad \ldots \\ &= k_0 + k_1 + \cdots\cdots + k_n \text{ if } c_n \leq x \leq b \end{aligned}$$

where $k_0, k_1, k_2, \ldots\ldots, k_n$ are all positive constants.

Here $c_1, c_2, \ldots\ldots, c_n$ are called step points. We observe that a step function f has a jump discontinuity at each step point, since at any such point c_i,

$\underset{x \to c_i + 0}{Lt} f(x) \neq \underset{x \to c_i - 0}{Lt} f(x)$ and the height of the jump at c_i is

$$f(c_i + 0) - f(c_i - 0) = \underset{x \to c_i + 0}{Lt} f(x) - \underset{x \to c_i - 0}{Lt} f(x) = k_i, \text{ for } i = 1, 2, \ldots\ldots n.$$

1.6. An Important Theorem, Some Important Principles and Inequalities in Classical Algebra.

Well Ordering Principle : Every non-empty subset of N has a least element, where N is the set of all natural numbers.

Archimedean Property of the natural order relation on the set R :

If a, b be two real numbers, where $a > 0$, then there exists a positive integer n such that $n\,a > b$.

Some important inequalities :

(*A*) If a, b be any two real numbers, then

(*i*) $|a \pm b| \leq |a| + |b|$

(*ii*) $|a \pm b| \geq ||a| - |b||$.

(*B*) If $a_1, a_2, \ldots\ldots, a_n ; b_1, b_2, \ldots\ldots b_n$ be real numbers, then

$$(a_1 b_1 + a_2 b_2 + \cdots\cdots + a_n b_n)^2 \leq (a_1^2 + a_2^2 + \cdots\cdots + a_n^2)(b^2 + b_2^2 + \cdots\cdots b_n^2).$$

This inequality is known as Cauchy's inequality.

(C) If $a_1, a_2, \ldots\ldots a_n$ be positive real numbers, then

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \geq \sqrt[n]{a_1 . a_2 \ldots . a_n} \geq \frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}$$

where equality sign occurs if and only if $a_1 = a_2 = \cdots = a_n$.

Multinomial Theorem :

If n be a positive integer and $a_1, a_2, \ldots\ldots, a_k$ are real numbers, then

$$(a_1 + a_2 + \cdots\cdots + a_k)^n = \sum_{i_1 + i_2 + \cdots + i_k = n} \frac{n!\, a_1^{i_1} a_2^{i_2} \ldots\ldots a_k^{i_k}}{i_1!\, i_2! \ldots\ldots i_k!}$$

where the summation includes all terms corresponding to the values of $i_1, i_2, \ldots\ldots, i_k$ belonging to $\{0, 1, 2\ldots, n\}$ subject to the condition

$$i_1 + i_2 + \cdots\cdots + i_k = n.$$

1.7. Euclidean Space R^n.

Let n be a given positive integer. We denote an ordered n-tuple of real numbers as a row matrix $(a_1, a_2, \ldots\ldots, a_n)$, where $a_i \in R$ for $i = 1, 2, \ldots\ldots, n$.

Let $R^n = \{(x_1, x_2, \ldots\ldots, x_n) : x_i \in R \text{ for } i = 1, 2, \ldots\ldots n\}$.

If $\bar{\alpha} = (x_1, x_2, \ldots \ldots, x_n)$ and $\bar{\beta} = (y_1, y_2, \ldots\ldots, y_n)$ be any two elements of R^n and x be any element of R, then $\bar{\alpha} + \bar{\beta}$, $x\bar{\alpha}$ are defined as

$$\bar{\alpha} + \bar{\beta} = (x_1 + y_1, x_2 + y_2, \ldots\ldots, x_n + y_n)$$

and $x\bar{\alpha} = (xx_1, xx_2, \ldots\ldots, xx_n)$.

These two operations are respectively called vector addition and scalar multiplication. Elements of R^n are called vectors and those of R will be called scalars.

Inner product of $\bar{\alpha}$ and $\bar{\beta}$, denoted as $\bar{\alpha}.\bar{\beta}$, is defined to be the real number $x_1 y_1 + x_2 y_2 + \cdots\cdots + x_n y_n$.

It can be shown that

(*i*) $\bar{\alpha} + \bar{\beta} = \bar{\beta} + \bar{\alpha}$ for all $\bar{\alpha}, \bar{\beta} \in R^n$.

(*ii*) $\bar{\alpha} + (\bar{\beta} + \bar{\gamma}) = (\bar{\alpha} + \bar{\beta}) + \bar{\gamma}$ for all $\bar{\alpha}, \bar{\beta}, \bar{\gamma} \in R^n$.

(*iii*) $\bar{\alpha} + \bar{o} = \bar{o} + \bar{\alpha} = \bar{\alpha}$ for all $\bar{\alpha} \in R^n$,
where $\bar{o} = (0, 0, \ldots\ldots, 0) \in R^n$

(*iv*) $\bar{\alpha} + (-\bar{\alpha}) = \bar{o}$, where $-\bar{\alpha} = (-1)\,\bar{\alpha}$, for all $\bar{\alpha} \in R^n$.

(*v*) $x(\bar{\alpha} + \bar{\beta}) = x\bar{\alpha} + x\bar{\beta}$
$(x + y)\bar{\alpha} = x\bar{\alpha} + y\bar{\alpha}$
$(xy)\,\bar{\alpha} = x\,(y\bar{\alpha})$
$1\bar{\alpha} = \bar{\alpha}$,
for all $\bar{\alpha}, \bar{\beta} \in R^n$ and for all $x, y \in R$.

(*vi*) $\bar{\alpha}.\bar{\beta} = \bar{\beta}.\bar{\alpha}$
$(\bar{\alpha} + \bar{\beta}).\bar{\gamma} = (\bar{\alpha}.\bar{\gamma}) + (\bar{\beta}.\bar{\gamma})$
$x(\bar{\alpha}.\bar{\beta}) = x\bar{\alpha}.\bar{\beta}$,
for all $\bar{\alpha}, \bar{\beta}, \bar{\gamma} \in R^n$, $x \in R$
and $\bar{\alpha}.\bar{\alpha} > 0$ if $\bar{\alpha} \neq \bar{o}$.

The set R^n with the above laws of 'vector addition', 'scalar multiplication' and 'inner product' is called n dimensional Euclidean space of ordered n-tuple of real numbers.

Linear dependence and independence :

Let $S = \{\bar{\alpha}_1, \bar{\alpha}_2, \ldots\ldots, \bar{\alpha}_k\}$ be a finite subset of R^n. S is said to be linearly dependent if there exist scalars $x_1, x_2 \ldots\ldots, x_k$, not all zero, such that $x_1 \bar{\alpha}_1 + x_2 \bar{\alpha}_2 + \cdots\cdots + x_k \bar{\alpha}_k = \bar{o}$, otherwise S is said to be linearly independent.

Basis of R^n :

Let $\bar{e}_i$ be an ordered n-tuple whose i-th component is 1 and the other components are 0, where i is a positive integer satisfying $1 \leq i \leq n$. It can be shown that (*i*) $\{\bar{e}_1, \bar{e}_2, \ldots\ldots, \bar{e}_n\}$ is linearly independent and (*ii*) any vector belonging to R^n can be expressed as $c_1\, \bar{e}_1 + c_2\, \bar{e}_2 + \cdots\cdots + c_n\, \bar{e}_n$ for suitable scalars $c_1, c_2, \ldots\ldots, c_n$.

A non-empty subset S of R^n will be called a basis of R^n if S has the properties (*i*) and (*ii*).

So $\{\bar{e}_1, \bar{e}_2, \ldots\ldots, \bar{e}_n\}$ is a basis of R^n.

Further it can be shown that a basis of R^n is not unique but the number of distinct vectors in every basis of R^n is n.

A basis $\{\bar{\alpha}_1, \bar{\alpha}_2, \ldots\ldots, \bar{\alpha}_n\}$ is called an orthonormal basis of R^n if

$$\bar{\alpha}_i . \bar{\alpha}_j = 0 \quad \text{for } i \neq j$$
$$= 1 \quad \text{for } i = j,$$

where $1 \leq i, j \leq n$.

Again a finite subset $\{\bar{\beta}_1, \bar{\beta}_2, \ldots\ldots, \bar{\beta}_k\}$ of R^n is called an orthonormal set if

$$\bar{\beta}_i . \bar{\beta}_j = 0 \quad \text{for } i \neq j$$
$$= 1 \quad \text{for } i = j$$

where $1 \leq i, j \leq k$.

We state (without proof) below an important theorem :

"An orthonormal set of vectors belonging to R^n can be extended to an orthonormal basis of R^n if the orthonormal set does not already form a basis of R^n."

We observe that the basis $\{\bar{e}_1, \bar{e}_2, \ldots\ldots, \bar{e}_n\}$ is an orthonormal basis of R^n.

From matrix algebra, we know that if A be an orthogonal matrix of size $n \times n$, then $A A^T = I$, where A^T is the transpose of A and I is the unit matrix of size $n \times n$. So we observe that if the set forms an orthonormal basis of R^n, then the vectors $\{\bar{\alpha}_1, \bar{\alpha}_2, \ldots\ldots, \bar{\alpha}_n\}$ belonging to the set form the rows of an orthogonal matrix.

1.8. Riemann-Stieltjes Integral with respect to a Monotonically Increasing Function.

Let g be a real valued monotonically increasing function defined on $[a, b]$. Also let f be a bounded real valued function on $[a, b]$. Let $P = \{x_0, x_1, \ldots\ldots, x_n\}$ be any partition of $[a, b]$, where $a = x_0 < x_1 < x_2 < \cdots\cdots < x_{n-1} < x_n = b$.

We write $\Delta g_i = g(x_i) - g(x_{i-1})$ for $i = 1, 2, \ldots\ldots n$.

Let $$U(P, f, g) = \sum_{i=1}^{n} M_i \, \Delta g_i,$$

$$L(P, f, g) = \sum_{i=1}^{n} m_i \, \Delta g_i,$$

where M_i, m_i are respectively the least upper bound and the greatest lower bound of f in $[x_{i-1}, x_i]$ for $i = 1, 2, \ldots\ldots n$.

It can be shown that both the sets

$A_1 = \{U(P, f, g) : P$ is a partition of $[a, b]\}$ and

$A_2 = \{L(P, f, g) : P$ is a partition of $[a, b]\}$

are bounded.

Then the greatest lower bound of A_1 exists and this greatest lower bound is denoted as $\overline{\int_a^b} f \, dg$ and the least upper bound of A_2 exists and this least upper bound is denoted by $\underline{\int_a^b} f \, dg$.

If $\underline{\int_a^b} f \, dg = \overline{\int_a^b} f \, dg$, we denote their common value by $\int_a^b f \, dg$ or sometimes by $\int_a^b f(x) \, dg(x)$, which is called the Riemann-Stieltjes integral of f with respect to g over $[a, b]$.

Some important properties of the Riemann-Stieltjes Integral :

1. If $\int_a^b f_1 dg$, $\int_a^b f_2 dg$ exist and c is a real constant, then

   (i) $\int_a^b (f_1 + f_2) dg = \int_a^b f_1 dg + \int_a^b f_2 dg,$

   (ii) $\int_a^b c \, f dg = c \int_a^b f dg.$

2. If $f_1(x) \leq f_2(x)$ on $[a, b]$, then

$$\int_a^b f_1 \, dg \leq \int_a^b f_2 \, dg,$$

provided the integrals exist.

3. If $\int_a^b f \, dg$ exists for some monotonically increasing function g on $[a, b]$, then

$$\left| \int_a^b f dg \right| \leq \int_a^b |f| \, dg.$$

We observe that if $g(x) = x$ and f is bounded on $[a, b]$, then $\int_a^b f \, dg$ is equal to Riemann integral $\int_a^b f \, dx$ provided it exists. If $\int_a^b f \, dx$ exists we say that f is bounded and integrable in $[a, b]$.

In this connection we state some important theorems.

1. If f is bounded and integrable on $[a, b]$ and if there exists a function ϕ such that $\frac{d}{dx} [\phi(x)] = f(x)$ on $[a, b]$, then

$$\int_a^b f \, dx = \phi(b) - \phi(a).$$

This is known as Fundamental Theorem of Integral Calculus.

2. Let f be bounded and integrable on $[a, b]$. A function $F : [a, b] \rightarrow R$ is defined by

$$F(x) = \int_a^x f(t) dt, \; a \leq x \leq b.$$

Then F is continuous on $[a, b]$ and furthermore if f is continuous at a point c of $[a, b]$, then $F'(c)=f(c)$.

3. If f is continuous on $[a, b]$, then f is integrable on $[a, b]$.

4. If f is bounded on $[a, b]$ and f has a finite number of points of jump discontinuity on $[a, b]$, then f is integrable on $[a, b]$.

We now state a theorem which gives a method of evaluating Riemann-Stieltjes integral by evaluating a Riemann integral.

If f is bounded and integrable on $[a, b]$ and g' is also integrable on $[a, b]$, then

$$\int_a^b f dg=\int_a^b f(x)\, g'(x)\, dx,$$

where g is a monotonically increasing function on $[a, b]$.

An important result on Riemann-Stieltjes integral:

Let $\{x_n\}$ be a sequence of distinct points in $(0, 1)$. Let $\sum_{n=1}^{\infty} c_n$ be convergent where $c_n > 0$ for all n.

Let $g(x)=\sum_{n=1}^{\infty} c_n\, \psi(x-x_n)$,

where $\psi(x)=0$, if $x \leq 0$
$\qquad\quad =1$, if $x > 0$.

Then, if f is continuous on $[0, 1]$, $\int_0^1 f dg=\sum_{n=1}^{\infty} c_n f(x_n)$.

1.9. Working Rule for Evaluating Double Integral of a Function $f(x, y)$ over a Region.

Here we consider double integral of a real valued bounded function $f(x, y)$ defined on a closed bounded subset D of $R \times R$.

If the double integral $\iint_D f\, dx\, dy$ exists for a bounded function defined on a closed domain D bounded by the curves

$$y=\phi(x),\ y=\psi(x)\ ;\ x=a,\ x=b,$$

where ϕ, ψ are continuous and $\phi(x) \leq \psi(x)$, for all $x \in [a, b]$ and if the integral $\int_{\phi(x)}^{\psi(x)} f\, dy$ exists for each $x \in [a, b]$, then the repeated integral

$$\int_a^b \left\{ \int_{\phi(x)}^{\psi(x)} f\, dy \right\} dx \text{ also exists and}$$

$$\iint_D f\, dx\, dy=\int_a^b \left\{ \int_{\phi(x)}^{\psi(x)} f\, dy \right\} dx.$$

In particular if D be the rectangular region $\{(x, y) : a \leq x \leq b,\ c \leq y \leq d\}$, then

$$\iint_D f\, dx\, dy=\int_a^b \left\{ \int_c^d f(x, y)\, dy \right\} dx$$

provided $\int_c^d f(x, y)\, dy$ exists for each $x \in [a, b]$.

Change of variable:

Let the double integral $\iint_D f(x, y)\, dx\, dy$ exist for a bounded function defined in a closed domain D of $R \times R$. Let $x=\phi(u, v)$, $y=\psi(u, v)$ represent a one-to-one mapping of the closed region D of xy plane into D' of the uv plane, where the functions $\phi(u, v)$ and $\psi(u, v)$ have continuous partial derivatives of first order at every point of D'.

If the Jacobian $J=\dfrac{\partial(x, y)}{\partial(u, v)} \neq 0$, then

$$\iint_D f(x, y)\, dx\, dy=\iint_{D'} f\{\phi(u, v), \psi(u, v)\}\, |J|\, du\, dv.$$

1.10. Improper Integrals, Beta Function and Gamma Function.

An integral $\int_a^b f(x)\, dx$ is said to be a proper integral if f is bounded in $[a, b]$ and a, b are finite. An integral is said to be an improper integral if it is not proper.

Let f be defined and let f be bounded in $[a+\epsilon, b]$ for every ϵ such that $0<\epsilon<b-a$ and let $\underset{x\to a+0}{Lt} f(x)=\infty$ or $\to -\infty$. Then the symbol $\int_a^b f(x)\,dx$ is said to be an improper integral where 'a' is a point of infinite discontinuity of f. If f is integrable in $[a+\epsilon, b]$ for every ϵ, where $0<\epsilon<b-a$ and if $\underset{\epsilon\to 0+}{Lt}\int_{a+\epsilon}^b f(x)\,dx$ exists finitely, then $\int_a^b f\,dx$ is said to be convergent and the value of this limit is called the value of the improper integral $\int_a^b f\,dx$. We can define similarly the convergence and the value of the improper integral $\int_a^b f\,dx$ where (*i*) f is defined in $[a, b]$ and f has an infinite discontinuity at b and f is bounded in $[a, b-\epsilon]$ for every ϵ where $0<\epsilon<b-a$, or, (*ii*) f is bounded in $[a, c-\epsilon_1]$ and $[c+\epsilon_2, b]$ for every ϵ_1, ϵ_2 where $a<c-\epsilon_1<c<c+\epsilon_2<b$ and f has an infinite discontinuity only at c where $a<c<b$. In case (*ii*), the value of $\int_a^b f\,dx$, when convergent, is given by

$$\int_a^b f\,dx = \underset{\substack{\epsilon_1\to 0+\\ \epsilon_2\to 0+}}{Lt}\left[\int_a^{c-\epsilon_1} f\,dx+\int_{c+\epsilon_2}^{b} f\,dx\right].$$

Improper integrals mentioned above are called improper integrals of type II. $\int_0^1 \frac{dx}{\sqrt{x}}$, $\int_{-1}^1 \frac{dx}{x^2}$, $\int_1^2 \frac{dx}{2-x}$ are improper integrals of type II of which only the first integral is convergent and

$$\int_0^1 \frac{dx}{\sqrt{x}} = \underset{\epsilon\to 0+}{Lt}\int_\epsilon^1 \frac{dx}{\sqrt{x}} = \underset{\epsilon\to 0+}{Lt}\left[2-2\sqrt{\epsilon}\,\right]=2.$$

The symbols $\int_{-x}^{x} f\,dx$, $\int_{-x}^{b} f\,dx$, $\int_{-x}^{x} f\,dx$ are called improper integrals of type I.

Let f be bounded and integrable in $[a, B]$ for every $B>a$. If $\underset{B\to\infty}{Lt}\int_a^B f\,dx$ exists finitely, then $\int_a^\infty f\,dx$ is said to be convergent and the value of this limit is called the value of the improper integral $\int_a^\infty f\,dx$. Similarly we can define the convergence and the value of $\int_{-\infty}^a f\,dx$.

Now let f be bounded and integrable in $[B_1, B_2]$ for all B_1, B_2 where $B_1<0$, $B_2>0$. If $\underset{\substack{B_1\to-\infty\\ B_2\to\infty}}{Lt}\int_{B_1}^{B_2} f\,dx$ exists finitely, then $\int_{-\infty}^{\infty} f\,dx$ is said to be convergent and the value of the limit is called the value of the improper integral $\int_{-\infty}^{\infty} f\,dx$.

$\int_{-\infty}^{-1}\frac{dx}{x^2}$, $\int_0^\infty e^{-x}\,dx$, $\int_{-\infty}^{\infty} e^{-x}\,dx$ are improper integrals of type I.

Let us examine the convergence of $\int_0^\infty e^{-x}\,dx$. We see that

$$\underset{B\to x}{LT}\int_0^B e^{-x}\,dx\ (B>0) = \underset{B\to x}{LT}\left(-e^{-B}+1\right)=1,\text{ which is finite.}$$

So $\int_0^\infty e^{-x}\,dx$ is convergent and its value is 1.

Finally, we introduce the concept of the convergence and the value of improper integral of the form $\int_a^\infty f\,dx$ where $\underset{x\to a}{Lt}|f(x)|=\infty$ and $c>a$. Such an integral is an improper integral of mixed type. $\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$ is an example of improper integral of mixed type and in this case we say that $\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$ is convergent if $\int_0^d \frac{e^{-x}}{\sqrt{x}}\,dx$ and $\int_d^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$ are both convergent where $d>0$ and the value of $\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$ is defined as $\int_0^d \frac{e^{-x}}{\sqrt{x}}\,dx+\int_d^\infty \frac{e^{-x}}{\sqrt{x}}\,dx$ provided the first improper integral of type II and the second improper integral of type I are both convergent.

Absolute Convergence :

The improper integrals $\int_a^\infty f\,dx$, $\int_{-\infty}^a f\,dx$, $\int_{-\infty}^\infty f\,dx$, $\int_a^b f\,dx$ are called absolutely convergent if $\int_a^\infty |f|\,dx$, $\int_{-\infty}^a |f|\,dx$, $\int_{-\infty}^\infty |f|\,dx$ and $\int_a^b |f|\,dx$ are respectively convergent.

A proper integral $\int_a^b f\,dx$ (when it exists) is also said to be convergent and it is said to be absolutely convergent if $\int_a^b |f|\,dx$ exists.

It can be shown that every absolutely convergent integral is convergent. Convergence and absolute convergence of improper integrals of the types $\int_{-\infty}^\infty f\,dg$ (Riemann-Stieltjes) and $\int_{-\infty}^\infty\int_{-\infty}^\infty f(x, y)\,dx\,dy$ can similarly be defined.

Beta Function :

It can be shown that the improper integral $\int_0^1 x^{m-1}(1-x)^{n-1}\,dx$ is convergent if and only if $m > 0, n > 0$. We denote the value of the integral (if convergent) by $B(m, n)$. The real valued function of two variables m, n defined by $B(m, n) = \int_0^1 x^{m-1}(1-x)^{n-1}\,dx$, if $m>0$, $n > 0$, is called the Beta Function.

Gamma Function :

It can be shown that the improper integral $\int_0^\infty e^{-x} x^{n-1}\,dx$ is convergent if and only if $n > 0$. We denote the value of the integral (if convergent) by $\Gamma(n)$. The real valued function of one real variable n, defined by $\Gamma(n) = \int_0^\infty e^{-x} x^{n-1}\,dx$, if $n > 0$, is called the Gamma Function.

Important properties and formulae.

1. $B(m, n) = B(n, m) = \int_0^\infty \dfrac{x^{m-1}}{(1+x)^{m+n}}\,dx$, if $m > 0, n > 0$.
2. $\Gamma(n+1) = n\Gamma(n)$, if $n > 0$.
   $\Gamma(n+1) = n!$, if n is a positive integer.
3. $\Gamma(1) = 1$.
4. $B(\frac{1}{2}, \frac{1}{2}) = \pi$.
5. $B(m, n) = \dfrac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$, if $m > 0, n > 0$.
6. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
7. $\int_0^\infty e^{-x^2}\,dx = \frac{1}{2}\Gamma(\frac{1}{2}) = \dfrac{\sqrt{\pi}}{2}$.
8. $$\int_0^{\pi/2} \sin^p x \cos^q x\,dx = \frac{1}{2}\,\frac{\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)}{\Gamma\left(\frac{p+q+2}{2}\right)},$$
   if $p > -1, q > -1$.

Examples I

1. Let $A = \{0, \frac{1}{2}, \frac{1}{4}, -1, 1\}$; $B = \{0, \frac{1}{2}\}$; $C = \{-1, 1, 4\}$; $D = \{-1\}$; $E = \{0, \frac{1}{4}\}$.

State which of the following statements are true : (Give reasons)

(i) $B \subset A$, (ii) $D \in A$, (iii) $B \cap D = \phi$,
(iv) $(B \cup C) \subset A$, (v) $A = B \cup C \cup D \cup E$,
(vi) Number of distinct elements of $B \times D$ is three.

2. In a class of 30 students, 15 students have taken English, 10 students have taken English but not Sanskrit. Find the number of students who have taken (i) Sanskrit and (ii) Sanskrit but not English. Assume that every student of the class takes at least one of lhe two subjects English and Sanskrit.

3. Show that the mapping $f: Z \to Z$ where $f(x) = |x|$ and $x \in Z$ is neither injective nor surjective where Z is the set of all integers.

4. Examine whether the mapping $f: R \to R$ has inverse where $f(x) = e^x$, $x \in R$ and R is the set of all real numbers.

5. Show that $\left\{\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right),\right.$

$$\left.\left(-\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)\right\}$$

is an orthonormal basis of R^3.

6. A function f is defined and bounded on $[-1, 1]$ and the function g is defined as follows :

$$\begin{aligned} g(x) &= 0 \text{ if } x < 0 \\ &= \tfrac{1}{2} \text{ if } x = 0 \\ &= 1 \text{ if } x > 0. \end{aligned}$$

If the Riemann-Stieltjes integral $\int_{-1}^{1} f\,dg$ exists then prove that its value is $f(0)$.

7. Evaluate :

(*i*) $\displaystyle\int_0^2 x\,dg$ where $g(x) = x$, if $0 \leqslant x \leqslant 1$

$= 2 + x$, if $1 < x \leqslant 2$.

(*ii*) $\displaystyle\int_0^4 x\,d\,([x] - x)$, where $[x]$ denotes the greatest integer not greater than x.

8. In a survey of 150 students, it was found that 40 students studied Economics, 50 students studied Mathematics, 60 students studied Accountancy and 15 students studied all the three subjects. It was also found that 27 students studied Economics and Accountancy, 35 students studied Accountancy and Mathematics and 25 students studied Economics and Mathematics. Find the number of students who studied only Economics and the number of students who studied none of these subjects.

Answers

1. (*i*) True, (*ii*) False, (*iii*) True, (*iv*) False, (*v*) False, (*vi*) False.

2. (*i*) 20, (*ii*) 15. 7. (*i*) 2, (ii) 2. 8. 3, 72.

CHAPTER II

THE CONCEPT OF PROBABILITY

2·1. Introduction.

The word 'probability' synonymous with the word 'chance' is used with reference to a class of experiments known as random experiments which will be defined precisely in the next section. We use the word 'probability' in many situations without conceiving a definite meaning of the term. For example, we say 'the probability that it will rain tomorrow is 70%', 'the probability of getting a head in tossing a coin is 40%', 'the probability of a new born baby to be a girl is 50%', etc. The word 'probability' used in the above statements expresses our degree of belief on happening of some events like 'it will rain tomorrow', 'we will get a head in tossing a coin', 'a new born baby will be a girl', etc. But actually here we try to characterise probability in an unscientific manner.

Before giving a proper meaning of the term 'probability' which can be used in science, we distinguish between two types of phenomena, namely (*i*) phenomena whose future behaviours are predictable in a deterministic manner, (*ii*) phenomena characterised by the fact that their future behaviour not predictable in a deterministic fashion. As an example of phenomena mentioned in (*i*), we consider the experiment of throwing a particle vertically upwards with a given initial velocity u under constant gravity g in a medium which offers no resistance. Here the greatest height H is determined uniquely by the well known formula $H = \frac{u^2}{2g}$. Here we can predict the value of the greatest height with certainty if the initial velocity u and the acceleration due to gravity g be known. Further if the experiment be repeated under identical conditions, the value of the greatest height will be same in every performance of the experiment. Now if we carefully examine the meaning of the term 'identical condition', we find that the experiment once performed cannot be performed again exactly under the same conditions, since in practice

5. Show that $\left\{\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)\right\}$

is an orthonormal basis of R^3.

6. A function f is defined and bounded on $[-1, 1]$ and the function g is defined as follows:

$$g(x) = 0 \text{ if } x < 0$$
$$= \tfrac{1}{2} \text{ if } x = 0$$
$$= 1 \text{ if } x > 0.$$

If the Riemann-Stieltjes integral $\int_{-1}^{1} f\,dg$ exists then prove that its value is $f(0)$.

7. Evaluate :

(i) $\int_0^2 x\,dg$ where $g(x) = x$, if $0 \leqslant x \leqslant 1$

$$= 2 + x, \text{ if } 1 < x \leqslant 2.$$

(ii) $\int_0^4 x\,d([x] - x)$, where $[x]$ denotes the greatest integer not greater than x.

8. In a survey of 150 students, it was found that 40 students studied Economics, 50 students studied Mathematics, 60 students studied Accountancy and 15 students studied all the three subjects. It was also found that 27 students studied Economics and Accountancy, 35 students studied Accountancy and Mathematics and 25 students studied Economics and Mathematics. Find the number of students who studied only Economics and the number of students who studied none of these subjects.

Answers

1. (i) True, (ii) False, (iii) True, (iv) False, (v) False, (vi) False.

2. (i) 20, (ii) 15. 7. (i) 2, (ii) 2, 8. 3, 72.

CHAPTER II

THE CONCEPT OF PROBABILITY

2·1. Introduction.

The word 'probability' synonymous with the word 'chance' is used with reference to a class of experiments known as random experiments which will be defined precisely in the next section. We use the word 'probability' in many situations without conceiving a definite meaning of the term. For example, we say 'the probability that it will rain tomorrow is 70%', 'the probability of getting a head in tossing a coin is 40%', 'the probability of a new born baby to be a girl is 50%', etc. The word 'probability' used in the above statements expresses our degree of belief on happening of some events like 'it will rain tomorrow', 'we will get a head in tossing a coin', 'a new born baby will be a girl', etc. But actually here we try to characterise probability in an unscientific manner.

Before giving a proper meaning of the term 'probability' which can be used in science, we distinguish between two types of phenomena, namely (*i*) phenomena whose future behaviours are predictable in a deterministic manner, (*ii*) phenomena characterised by the fact that their future behaviour not predictable in a deterministic fashion. As an example of phenomena mentioned in (*i*), we consider the experiment of throwing a particle vertically upwards with a given initial velocity u under constant gravity g in a medium which offers no resistance. Here the greatest height H is determined uniquely by the well known formula $H = \frac{u^2}{2g}$. Here we can predict the value of the greatest height with certainty if the initial velocity u and the acceleration due to gravity g be known. Further if the experiment be repeated under identical conditions, the value of the greatest height will be same in every performance of the experiment. Now if we carefully examine the meaning of the term 'identical condition', we find that the experiment once performed cannot be performed again exactly under the same conditions, since in practice

the values of the initial velocity u in any two trials (a particular performance of the experiment will be called a trial of the experiment) cannot be made exactly equal. But the variations in the values of u in different trials are very small so that we can neglect such variations and therefore, in spite of such variations we state that the experiment is repeated under indentical conditions. So 'identical conditions' mean not exactly identical but as identical as possible. Thus in practice, the truth of the statement 'greatest height is same in every performance of the experiment under identical conditions' can be verified only in the above approximate sense and in this sense we can state 'identical values of u under identical given conditions yield identical values of greatest height.' This is true for every such experiment. So the basic conviction of science identical conditions yield identical results' can be verified only in the approximate sense.

From now the phrase 'identical conditions' or 'uniform conditions' will always be understood in this approximate sense.

The experiment mentioned above belongs to a class of experiments for which we can state 'identical conditions yield identical results' and so results are predictable in a deterministic fashion.

As an example of the phenomena mentioned in (*ii*) we consider the experiment of tossing a coin. If we repeat the experiment under identical conditions, it will be observed that the possible results will be either a 'head' or a 'tail' and an exact prediction of the result in any particular trial is always impossible. Similar remarks can be made for other experiments of this type, like throwing a die, drawing a card from a pack of cards, predicting the sex of a new born baby, telling the number of particles emitted by a radioactive source in a given interval of time, etc. We do not attempt to explain the reasons for unpredictability of the results with reference to such experiments but we want to give emphasis on the fact that if such an experiment be repeated under identical conditions the results are not uniquely determined by the initial conditions but vary at random.

Thus we see that there is a class of experiments which are such that exact prediction of the results of individual trials are

impossible. But if we turn our attention from the individual trials to the whole sequence of trials of a given experiment of the above class, we shall see that the results of a long sequence of trials of a given experiment show a striking regularity. This type of regularity is known as 'statistical regularity' and it will be explained in § 2.7. On the basis of this regularity a definite and concrete meaning of the statement 'the probability that it will rain tomorrow in a locality is 70% can be given as follows : 'If we conceive of a large number of days in the past with conditions like today, then approximately 70 per cent of the days were followed by days with rain, 30 per cent were followed by days without rain.

The above mentioned regularity is observed in many natural phenomena and so application of probability theory to Physics, Biology, Engineering, Economics, Statistics or any other branch of Science, actually proceeds from the conviction in the existence of probabilities based on statistical regularity. We give an example of the application of probability in Physics based on this conviction. From the point of view of Molecular Physics, every substance consists of an enormous number of small particles, in constant interaction. Little is known about the nature of these particles, their interaction, mode of motion, etc. But the problem here is not to study the individual particle motion but to investigate the regularity that arises in assemblies of large number of moving and interacting particles. Thus in 1802, Dalton enunciated his law for the pressure of gas mixtures, was based on the tacit assumption that the motion of all the particles involved as uniform. By the middle of the century Clausius (1847) and Joule (1857) had shown how to express the pressure in terms of the mean velocity of the gas molecules. By 1860 Maxwell applied these ideas to the random motions of gas molecules and from this Statistical theory of gases was rapidly developed.

At the end, we mention some important applications of theory of probability in various branches of knowledge. Gauss (1777-1855) and Laplace (1749-1827) discussed independently the applications of the theory of probability to numerical analysis of errors of measurements in physical and astronomical observations. The

enormous development of life insurance since the beginning of the nineteenth century was rendered possible by a corresponding development of actuarial mathematics, which is based on the application of probability to mortality statistics. Methods of mathematical statistics have been introduced into many fields of practical and scientific activities. The mathematical theory on which these methods are based rests on the foundations of the theory of probability.

The theory of probability, which at the present day is an important branch of mathematics, with a wide field of applications as mentioned above, has developed from a very abased origin. The true origin of the theory lies in the correspondence between two great men of the seventeenth century, Pascal (1623-1662) and Fermat (1601-1665). In the French society of the 1650's, gambling was a popular and fashionable habit. With the introduction of more complicated games with cards, dice, coins, etc. where enormous sums of money were at stake in gambling establishments, the need was felt for a logical method for calculating the chances of winning for gamblers in various games. A French nobleman Chevelier De Méré, a man of ability and great experience in gambling had the idea of consulting the famous mathematician and philosopher Pascal in Paris on some questions with certain games of chance and this gave rise to a correspondence between Pascal and Fermat. This correspondence forms the origin of probability theory. At this early phase of the development no systematic theory of probability had been worked out, and the whole subject consisted of a collection of isolated problems concerning various games. At this stage, the basic concepts were not defined with sufficient precision. This vagueness frequently led to paradoxical conclusions (*e.g.*, Bertrand's paradox—Gnedenko p-34, Ex. 2). So it became necessary to study systematically the basic concepts of probability theory and to clarify the conditions under which the results of the theory could be employed.

In the next sections of this chapter we shall define and explain the basic concepts like random experiments, events, statistical regularity etc. and discuss various attempts made to define probability leading to a rigorous foundation of the mathematical theory of probability.

2.2. Random Experiment.

An experiment is generally thought of as one or more acts which result in some outcome.

An experiment E is called a random experiment if (*i*) all possible outcomes of E are known in advance, (*ii*) it is impossible to predict which outcome will occur at a particular performance of E, (*iii*) E can be repeated, at least conceptually, under identical conditions for infinite number of times.

The experiment of tossing a coin is an example of random experiment. Here the possible outcomes are 'head' and 'tail', but it is impossible to predict which outcome, namely 'head' or 'tail', will occur at a particular toss of the coin under the given conditions.

Other examples of random experiment are 'throwing a die', 'drawing a card from a full pack of 52 cards at random', etc.

We observe that outcomes are not uniquely determined by a given random experiment. The outcomes are determined by the purpose for which the experiment is carried out.

Let us consider the random experiment of noting whether two given components of a machine are functioning properly or not. If our purpose is to count only the number of components functioning properly (but not interested in exactly which of the components are functioning properly) then there are only three possible outcomes, namely, 'two functioning properly', 'two malfunctioning', 'one functioning properly and one malfunctioning'. But if our purpose is to note exactly which of the components are functioning properly, then denoting two given components as first component and second component, we see that the possible outcomes are 'first functioning properly, second functioning properly'; 'first functioning properly, second malfunctioning'; 'first malfunctioning, second functioning properly'; 'first malfunctioning. second malfunctioning'; *i.e.*, in this case there are four outcomes.

2.3. Event Space.

The set of all possible outcomes (determined for a given purpose) of a given random experiment E is called the event space of the experiment and it will be denoted by S. Here the outcomes, also called event points, are the elements of S. The event space S of the random experiment of tossing a coin is $\{H, T\}$, where H denotes the outcome 'head' and T denotes the outcome 'tail'. This is an example of a finite event space. The event space S corresponding to the experiment of choosing a number at random from the interval $(2, 4)$ is the set $(2, 4)$ which is an infinite set. Another example of finite event space is given below :

Let E be the random experiment of throwing a pair of dice. The corresponding event space S is given by

$$S=\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), \ldots, (6, 6)\}$$

which contains 36 distinct outcomes.

2.4. Events.

Intuitively an event of a given random experiment can be looked upon as a statement whose truth or falsity is determined after the experiment. Let P be a statement and let A be the set of all outcomes (of E) for which P is true. Then the event expressed by P can be described by A which is a subset of S, where S is the event space of E and we say that A is an event. We know that only one outcome belonging to S will occur in a particular performance of E. Now if 'a' be any element of A and 'a' occurs at a specific trial of E, we say that the event A has happened and if an outcome 'b' occurs where $b \notin A$, we say that A has not happened.

Let us consider the random experiment of throwing a die. Here $S=\{1, 2, 3, 4, 5, 6\}$. Let $A=\{2, 4, 6\}$ be an event which can be described verbally as 'even number appears in throwing a die'. Here the event A happens in a specific trial of the given random experiment if and only if exactly one of the outcomes '2', '4' or '6' occurs in the trial.

From above we find that formally an event A of a given random experiment can be defined as a subset of the corresponding event space S. To avoid certain difficulties, we must place restrictions

on subsets (of S), termed as events, to which probabilities can be assigned (to be defined later). In a given problem there will be a particular class Δ of subsets of S such that any member of Δ can be called an event and Δ will be called the class of events. If the event space S be at most countable, then every subset of S can be an event. But this is not true when S is uncountable. Let $S=[0, 1]$ which is uncountable and let the probability assigned be such that $P(A)=b-a$, where A denotes the event $\{x : a \leq x \leq b\}$ and $0 \leq a \leq b \leq 1$. Then it can be shown that not all possible subsets of S can be assigned probabilities in a manner consistent with the axioms of probability stated in § 3.1.

Before defining precisely which subsets of S can be called events, we shall describe some particular events of a random experiment verbally and give the corresponding set theoretic notations.

Let A and B be any two events of a given random experiment. The union of the sets A and B, denoted by $A+B$, is an event which can be verbally stated as 'A or B' or equivalently 'at least one of A and B'. The intersection of A and B, denoted as AB, is an event which can be verbally stated as 'both A and B'. The complement of A, denoted as $\bar{A}$, is an event verbally stated as 'not A'. In general, the union and intersection of finite or countably infinite number of events of a given random experiment are also events of the same random experiment. If $A_1, A_2, \ldots, A_n, \ldots$ are events, then the event $\sum_{n=1}^{\infty} A_n$ is verbally stated as 'at least one of $A_1, A_2, \ldots, A_n, \ldots$' and the event $\prod_{n=1}^{\infty} A_n$ is verbally stated as 'all of $A_1, A_2, \ldots, A_n, \ldots$'.

Impossible event : An event of a given random experiment is called an impossible event if it can never happen in any performance of the random experiment under identical conditions. Such an event is described by the empty subset O of the corresponding event space S. In connection with the random experiment of throwing a die, the event 'face marked 7' is an impossible event.

Certain event : An event of a given random experiment is called 'certain event', if it happens in every performance of the

corresponding random experiment under identical conditions. Formally a certain event is described by the set S, which is the corresponding event space. In connection with the random experiment of tossing a coin, the event 'head or tail' is a certain event and it is described by the event space $S=\{H, T\}$.

Now we shall not give details of mathematics to precisely explain why the above mentioned subsets of S should be called events. Instead we state the following properties which seem reasonable to define the class Δ of subsets of S forming the class of events of a given random experiment :

(*i*) $S \in \Delta$.

(*ii*) If $A \in \Delta$ then $\bar{A} \in \Delta$.

(*iii*) If $A_1, A_2, \ldots\ldots A_n, \ldots\ldots \in \Delta$ then

$$\sum_{n=1}^{\infty} A_n \in \Delta.$$

Any member of Δ will be called an event of the given random experiment and this is consistent with the notion of events introduced at the beginning of this section.

A class Δ of subsets of a given set S, satisfying (*i*), (*ii*), (*iii*) stated above is called a σ-algebra or a σ field or a Borel field.

The following theorems follow from the properties (*i*), (*ii*), (*iii*) of Δ.

THEOREM 2.4.1. $O \in \Delta$.

Proof: By (*i*) $S \in \Delta$

Then by (*ii*) $\bar{S} \in \Delta$. But $\bar{S}=O$.

Hence $O \in \Delta$.

THEOREM 2.4.2. If $A_1, A_2, \ldots\ldots, A_n \in \Delta$, then

$$A_1+A_2+\cdots\cdots+A_n \in \Delta$$

and $A_1A_2\ldots\ldots A_n \in \Delta$.

Proof: By Theorem 2.4.1, we see that $O \in \Delta$.

Then if we take $A_i=O$ for $i=n+1, n+2, \ldots\ldots$

then $A_1, A_2, \ldots\ldots, A_n, A_{n+1}, \ldots\ldots \in \Delta$.

By (*iii*), we get

$$A_1+A_2+\cdots\cdots+A_n+O+O+\cdots\cdots \in \Delta.$$

Hence $A_1+A_2+A_3+\cdots\cdots+A_n \in \Delta$.

Again by (*ii*) $\bar{A}_1, \bar{A}_2, \ldots\ldots, \bar{A}_n \in \Delta$.

So $\bar{A}_1+\bar{A}_2+\cdots\cdots+\bar{A}_n \in \Delta$, by the first part of the theorem.

Then by (*ii*) $\overline{\bar{A}_1+\bar{A}_2+\cdots\cdots+\bar{A}_n} \in \Delta$.

Now by De Morgan's law,

$$\overline{\bar{A}_1+\bar{A}_2+\cdots\cdots\bar{A}_n}=\bar{\bar{A}}_1\,\bar{\bar{A}}_2\ldots\ldots\bar{\bar{A}}_n$$
$$=A_1\,A_2\ldots\ldots A_n.$$

Hence, $A_1\,A_2\ldots\ldots A_n \in \Delta$.

THEOREM 2.4.3. If $A_1, A_2, \ldots\ldots, A_n, \ldots\ldots \in \Delta$,

then $\prod_{n=1}^{\infty} A_n \in \Delta$.

Proof: By (*ii*) and (*iii*), we get $\sum_{n=1}^{\infty} \bar{A}_n \in \Delta$. Now by De Morgan's law, $\overline{\sum_{n=1}^{\infty} \bar{A}_n} = \prod_{n=1}^{\infty} A_n$. Now by (*ii*) $\overline{\sum_{n=1}^{\infty} \bar{A}_n} \in \Delta$.

Hence $\prod_{n=1}^{\infty} A_n \in \Delta$.

N. B. The impossible event O never happens in any performance of the corresponding random experiment E and the certain event S happens in every performance of E, but any event $A(\neq O, S)$ may or may not happen when E is performed under a given set of conditions and for this reason an event other than impossible event and certain event is called a random event. But we shall use 'random event' to mean any event (impossible or certain or any other event) of a random experiment and as such 'events' will also be called 'random events'.

2.5. Simple and Composite Events.

An event A is called a simple event or an elementary event if A contains exactly one element, *i.e.*, A can happen in only one way in any performance of the corresponding random experiment.

An event is called a composite event if A contains more than one element. Some authors use 'compound event' in place of 'composite event'.

In connection with the random experiment of throwing an ordinary die, the event space S is given by

$$S=\{1, 2, 3, 4, 5, 6\}.$$

Let A_1, B_1, C_1 be respectively the three events defined by

$$A_1=\{2, 4, 6\}, \quad B_1=\{3, 6\}, \quad C_1=\{2\}.$$

The events A_1 and B_1 are composite events whereas the event C_1 is a simple event.

N. B. Many authors use 'Sample space' in place of 'Event space'. But we shall use 'Event space' throughout the present treatise.

2.6. Mutually Exclusive Events.

Two events connected to a given random experiment E are said to be mutually exclusive if A, B can never happen simultaneously in any performance of E, *i.e.*, if $AB=O$. In connection with the random experiment of throwing a die, the events 'multiple of 3' and 'a prime number' are not mutually exclusive, since the number '3' is a multiple of 3 as well as a prime number, whereas the events 'even number' and 'odd number' are mutually exclusive events of the same random experiment.

2.7. Exhaustive Set of Events.

A collection of events is said to be exhaustive if in every performance of the corresponding random experiment at least one event (not necessarily the same for every performance) belonging to the collection happens. In set theoretic notations the collection of events $\{A_\alpha : \alpha \in I\}$ is exhaustive if and only if

$$\sum_{\alpha \in I} A_\alpha = S,$$

where I is an index set and S is the corresponding event space. In connection with the random experiment of throwing a die the collection of events $\{A_1, A_2, A_3\}$ is exhaustive where

$$A_1=\{1, 3, 5\}, \quad A_2=\{2\}, \quad A_3=\{4, 6\}.$$

2.8. Statistical Regularity.

Let a random experiment E be repeated N times under identical conditions, in which we note that an event A of E occurs $N(A)$

times. Then the ratio $\frac{N(A)}{N}$ is called the frequency ratio of A and is denoted by $f(A)$. Now if the random experiment E is repeated a very large number of times, it is seen that the frequency ratio $f(A)$ gradually stabilises to a more or less constant, *i.e.*, $f(A)=\frac{N(A)}{N}$ gradually tends to a constant number as N becomes larger and larger. This tendency of stability of frequency ratio is called statistical regularity and this fact was confirmed by many experimental results.

2.9. Classical Definition of Probability.

Towards the beginning of the 19th century, Laplace gave a formal definition of probability which goes by the name of the classical definition. The theory of probability developed on the basis of the classical definition is known as the classical theory of probability. In the classical theory we have the following definition of probability :

Let the event space S of a given random experiment E be finite. If all the simple events connected to E be 'equally likely' then the probability of an event $A (A \subseteq S)$ is defined as

$$P(A)=\frac{m}{n},$$

where n is the total number of simple events connected to E, *i.e.*, n is the number of distinct elements of S and m of these simple events are favourable to A, *i.e.*, A contains m distinct elements.

At this stage it is not possible to give a precise definition of the phrase 'equally likely' used in the above definition. In the next section we shall critically examine the meaning of the phrase. At present, we shall say that all the simple events are equally likely if it is understood intuitively that no one of them is expected to occur in preference to others in any trial of the given random experiment and only then the definition can be applied.

Deductions :—

(*a*) $0 \le P(A) \le 1$

(*b*) $P(S)=1$

(c) $P(O)=0$

(d) $P(\bar{A})=1-P(A)$.

Proof: (a) We have $P(A)=\frac{m}{n}$, where m, n have the meanings given before.

Here $0 \leq m \leq n$

or, $0 \leq \frac{m}{n} \leq 1$.

So, $0 \leq P(A) \leq 1$.

(b) $P(S)=\frac{n}{n}=1$.

(c) $P(O)=\frac{0}{n}=0$.

(d) $P(\bar{A})=\frac{n-m}{n}=1-\frac{m}{n}=1-P(A)$.

Theorem 2.9.1. Theorem of Total Probability.

If $A_1, A_2, \ldots, A_k$ are pairwise mutually exclusive events, then

$P(A_1+A_2+\cdots+A_k)=P(A_1)+P(A_2)+\cdots+P(A_k)$.

Proof: Let n be the total number of simple events of the corresponding random experiment E of which m_i are favourable to A_i, $i=1, 2, \ldots, k$. Since the events $A_1, A_2, \ldots, A_k$ are mutually exclusive, the total number of simple events favourable to the event $A_1+A_2+\cdots+A_k$ is $m_1+m_2+\cdots+m_k$,

$0 \leq m_i \leq n$, $i=1, 2, \ldots, k$.

Then by the classical definition,

$$P(A_1+A_2+\cdots+A_k)=\frac{m_1+m_2+\cdots+m_k}{n}$$
$$=\frac{m_1}{n}+\frac{m_2}{n}+\cdots+\frac{m_k}{n}$$
$$=P(A_1)+P(A_2)+\cdots+P(A_k).$$

Hence the theorem.

2.10. Criticisms of the Classical Definition.

If we examine the classical definition a little more closely we find that there is a logical drawback in the definition. We note that the definition can be used only if it is possible to ascertain that all the simple events are equally likely. In many problems, considerations of symmetry and similarity enable us to decide whether, in the problem before us, simple events are equally likely. For example, if a die be symmetric, then the simple events connected to the random experiment of throwing the die may be considered to be equally likely. But it is very difficult to explain the nature of 'symmetry' and 'similarity' as stated above. It was found after many serious investigations that the phrase 'equally likely' cannot be explained without the prior idea of probability.

Moreover, the definition is restricted to event spaces which are finite and where all the simple events are equally likely. The definition cannot be applied where the simple events are not equally likely or where the event space is infinite. With the help of this definition it will thus be impossible to treat the case of a loaded die since here intuitively we can expect that a face can turn up in preference to others and consequently simple events are not necessarily equally likely and the case of predicting the number of telephone calls in a given interval (in a given trunk line) in which there are infinite number of simple events.

In order to avoid the limitations of the classical approach and to make the definition more widely applicable, we now take recourse in the next section to another definition, called the frequency definition of probability.

It may be noted here that the classical definition is based on advance subjective concept of probability so that $P(A)=\frac{m}{n}$ should rather be called a method of calculation of probability for events of a finite event space of equally likely simple events, instead of taking it as a definition of probability.

2.11. Frequency Definition of Probability.

Let A be an event of a given random experiment E. Let the event A occur $N(A)$ times when the random experiment E is repeated N times under identical conditions. Then on the basis of statistical regularity we can assume that $\underset{N\to\infty}{Lt} \frac{N(A)}{N}$ exists finitely

and the value of this limit is called the probability of the event A, denoted by $P(A)$,

$$i.e., \ P(A) = \underset{N \to \infty}{Lt} \frac{N(A)}{N} = \underset{N \to \infty}{Lt} f(A),$$

where $f(A) = \frac{N(A)}{N}$ is the frequency ratio of the event A in N repetitions of the corresponding random experiment under identical conditions.

Deductions :

(a) $0 \leq P(A) \leq 1$, for any event A

(b) $P(S) = 1$

(c) $P(O) = 0$

(d) $P(\bar{A}) = 1 - P(A)$.

Proof : (a) We have $0 \leq N(A) \leq N$, where N and $N(A)$ have the meanings given above.

$$\therefore \ 0 \leq \frac{N(A)}{N} \leq 1$$

$$\text{or, } 0 \leq \underset{N \to \infty}{Lt} \frac{N(A)}{N} \leq 1.$$

Hence, $0 \leq P(A) \leq 1$.

(b) $$P(S) = \underset{N \to \infty}{Lt} \frac{N(S)}{N} = \underset{N \to \infty}{Lt} \frac{N}{N} = 1.$$

(c) $$P(O) = \underset{N \to \infty}{Lt} \frac{N(O)}{N} = \underset{N \to \infty}{Lt} \frac{0}{N} = 0.$$

(d) $$P(\bar{A}) = \underset{N \to \infty}{Lt} \frac{N(\bar{A})}{N}$$

$$= \underset{N \to \infty}{Lt} \frac{N - N(A)}{N}$$

$$= \underset{N \to \infty}{Lt} \left[1 - \frac{N(A)}{N}\right]$$

$$= 1 - \underset{N \to \infty}{Lt} \frac{N(A)}{N}$$

$$= 1 - P(A).$$

Theorem 2.11.1. Theorem of Total Probability.

If $A_1, A_2, \ldots\ldots, A_n$ be n pairwise mutually exclusive events, then

$$P(A_1 + A_2 + \cdots\cdots + A_n) = P(A_1) + P(A_2) + \cdots\cdots + P(A_n).$$

Proof : We give a proof of the theorem applying frequency definition.

We have, $$P(A_1 + A_2 + \cdots\cdots + A_n) = \underset{N \to \infty}{Lt} \frac{N(A_1 + A_2 + \cdots + A_n)}{N}$$

Now $$N(A_1 + A_2 + \cdots\cdots + A_n) = \sum_{i=1}^{n} N(A_i),$$

since $A_i A_j = O$, whenever $i \neq j$, the events being pairwise mutually exclusive.

$$\therefore \ P(A_1 + A_2 + \cdots + A_n) = \underset{N \to \infty}{Lt} \frac{N(A_1) + N(A_2) + \cdots\cdots + N(A_n)}{N}$$

$$= \underset{N \to \infty}{Lt} \left[\frac{N(A_1)}{N} + \frac{N(A_2)}{N} + \cdots\cdots + \frac{N(A_n)}{N}\right]$$

$$= \underset{N \to \infty}{Lt} \frac{N(A_1)}{N} + \underset{N \to \infty}{Lt} \frac{N(A_2)}{N} + \cdots\cdots + \underset{N \to \infty}{Lt} \frac{N(A_n)}{N}$$

$$= P(A_1) + P(A_2) + \cdots\cdots + P(A_n).$$

Hence the theorem.

Deduction of Classical Definition

In this case the event space S is finite and contains n distinct elements $u_1, u_2, \cdots, u_n$ (say), so that

$$S = \{u_1, u_2, \ldots, u_n\}.$$

Here the n distinct simple events $U_1 = \{u_1\}$, $U_2 = \{u_2\}, \ldots$, $U_n = \{u_n\}$ are equally likely, which means that the simple events have equal probability, *i.e.*, $P(U_1) = P(U_2) = \cdots = P(U_n)$. Since any two distinct simple events are necessarily mutually exclusive, we have

$$1 = P(S) = P(U_1 + U_2 + \cdots + U_n) = P(U_1) + P(U_2) + \cdots + P(U_n).$$

$$\therefore \ P(U_1) = P(U_2) = \cdots = P(U_n) = \frac{P(U_1) + P(U_2) + \cdots + P(U_n)}{n} = \frac{1}{n}.$$

Let now A be any event connected to the given random experiment. If the event A contains m distinct elements of S, say, $u_{i_1}, u_{i_2}, \ldots\ldots, u_{i_m}$, where $i_1, i_2, \cdots\cdots, i_m$ take distinct values from the set $\{1, 2, \cdots\cdots, n\}$, we can write $A = U_{i_1} + U_{i_2} + \cdots\cdots + U_{i_m}$.

Now let E be repeated under identical conditions N times in which A occurs $N(A)$ times.

Also let U_{i_1} occur k_1 times,

U_{i_2} occur k_2 times,

... ...

U_{i_m} occur k_m times.

Then $N(A) = k_1 + k_2 + \cdots \cdots + k_m$, since U_{i_1}'s, $U_{i_2}, \cdots \cdots, U_{i_m}$ are pairwise mutually exclusive events.

$$\therefore \quad \frac{N(A)}{N} = \frac{k_1 + k_2 + \cdots\cdots + k_m}{N} = \frac{N(U_{i_1}) + N(U_{i_2}) + \cdots\cdots + N(U_{i_m})}{N}$$

$$\therefore \quad \underset{N\to\infty}{Lt} \frac{N(A)}{N} = \underset{N\to\infty}{Lt} \frac{N(U_{i_1}) + N(U_{i_2}) + \cdots\cdots + N(U_{i_m})}{N}$$

$$= \underset{N\to\infty}{Lt} \left[\frac{N(U_{i_1})}{N} + \frac{N(U_{i_2})}{N} + \cdots\cdots + \frac{N(U_{i_m})}{N}\right].$$

So, applying frequency definition of probability,

$$P(A) = P(U_{i_1}) + P(U_{i_2}) + \cdots\cdots + P(U_{i_m}).$$

But $P(U_{i_1}) = P(U_{i_2}) = \cdots\cdots = P(U_{i_m}) = \dfrac{1}{n}$.

$\therefore \quad P(A) = \dfrac{m}{n}$, which gives the classical definition of probability.

2.12. Conditional Probability.

Let E be a given random experiment and A be an event of E. We have discussed methods for computing the probability $P(A)$ on the basis of only information that any outcome of E can occur in a trial of E. Now suppose we are given the added information that an outcome u of a trial is contained in a subset B of the event space of E, *i.e.*, it is given that the event B has happened in a trial of E. Knowledge of the occurrence of the event B may change the probability of the event A. We wish to define the probability of the event A, given that the event B occurs. Let us give examples from the real life situations where such conditional probabilities occur. In the experiment of finding the life of a light-bulb, we might be interested in the probability that the bulb will last 75 hours, given that it has already lasted 20 hours. In the experiment of throwing a die one might be interested in the probability of the event 'multiple of 3', given that the event 'even number' occurs in a particular throwing. Probability questions of the above type are considered in the framework of conditional probability. Keeping in mind the above notion, we shall define conditional probability.

Definition of conditional probability.

(Let E be a given random experiment and A, B be two events of E where $P(B) \neq 0$. The conditional probability of the event A on the hypothesis that the event B has happened, denoted by $P(A \mid B)$, is defined by

$$P(A \mid B) = \underset{N\to\infty}{Lt} \frac{N(AB)}{N(B)},$$

assuming that the limit exists, N, $N(AB)$, $N(B)$ have the usual meanings given before.)

Theorem 2.12.1. Theorem of Compound Probability.

If A, B are two events of a given random experiment, then

$$P(AB) = P(A/B)\, P(B), \text{ if } P(B) \neq 0$$

$$\text{or, } P(AB) = P(B/A)\, P(A), \text{ if } P(A) \neq 0.$$

Proof: Let E be the given random experiment and let E be repeated under identical conditions N times. If $N(AB)$, $N(B)$, $N(A)$ be respectively the number of occurrences of the events AB, B, A, then

$$P(A/B) = \underset{N\to\infty}{Lt} \frac{N(AB)}{N(B)} \quad [\text{ Here we note that } P(B) > 0 \Rightarrow N(B) > 0 \text{ for suitable large } N.]$$

$$= \underset{N\to\infty}{Lt} \frac{\frac{N(AB)}{N}}{\frac{N(B)}{N}}$$

$$\text{or,}\quad P(A/B)=\frac{\underset{N\to\infty}{Lt}\dfrac{N(AB)}{N}}{\underset{N\to\infty}{Lt}\dfrac{N(B)}{N}}$$

$$=\frac{P(AB)}{P(B)}.\quad\left(\because\ \underset{N\to\infty}{Lt}\frac{N(B)}{N}=P(B)\neq 0\right).$$

Hence we get.

$$P(AB)=P(A\mid B)\,P(B)\ \text{if}\ P(B)\neq 0.$$

It can be proved similarly that

$$P(AB)=P(B\mid A)\,P(A)\ \text{if}\ P(A)\neq 0.$$

Remark: We know that the event S occurs in any trial of a given random experiment, so from the meaning of conditional probability given above we find that unconditional probability $P(A)$ is a particular conditional probability, since the statement 'A occurs' can also be expressed as 'A occurs on the hypothesis that S has happened'. So $P(A\mid S)$ and $P(A)$ should be equal and by the theorem of compound probability we find that

$$P(A\mid S)=\frac{P(AS)}{P(S)}=\frac{P(A)}{1}=P(A).$$

2·13. Criticisms of the Frequency Definition.

In this definition we note that the frequency ratio $\frac{N(A)}{N}$ is obtained from observation whereas $\underset{N\to\infty}{Lt}\frac{N(A)}{N}$ is a rigorous analytical concept. This combination of empirical and analytical concepts leads to mathematical difficulties. Although there is not much objection against the logical content of the theory of probability based on the frequency definition but due to the aforesaid weakness in the definition it will be unwise to build the theory of probability on the basis of this definition.

Conclusion: In sections 2.10 and 2.13 we have seen that classical and frequency definitions are both inadequate for developing the mathematical theory of probability. Now the theory of probability is conceived as a mathematical theory of phenomena showing statistical regularity. So in order that mathematical theory

of probability may be applied to different types of phenomena showing statistical regularity, the definition of probability should be independent of the intended application. From all these considerations we feel the necessity of an axiomatic treatment of the theory of probability, *i.e.,* the theory of probability, as a branch of mathematics, should be developed from axioms in exactly the same way as Geometry and Algebra. Axioms are propositions which are regarded as true and not proved within the framework of the given theory. All other propositions of the theory have to be proved from the accepted axioms in a purely logical manner. Axiomatic theory starts from one or more sets of abstract objects, where some relations between the objects are expressed by the axioms. The points, lines, planes considered in Pure Geometry as abstract objects, are not things that we know from immediate experience. Pure Geometry deals with such abstract objects entirely defined by their properties, as expressed by the five sets of axioms, namely, 'axioms of incidence', 'axioms of order', 'axioms of motion', 'axiom of parallelism' and 'axiom of continuity' (Hilbert).

Now any mathematical theory developed logically from a set of axioms can have many concrete interpretations besides those from which the axioms are developed. Similar is the situation in the axiomatic theory of probability. But we shall interpret the theory in such a way that 'events' will be events of the real world and the probability will be so interpreted that it can be applied to phenomena showing statistical regularity.

Formulation of the axioms are the results of a prolonged accumulation of facts and a logical analysis of the results obtained and in this way the axioms of Geometry, studied in elementary mathematics, were formulated. The axioms taken for defining probability will certainly be motivated by the results obtained from the classical and frequency definition of probability. On the basis of the axioms it will be possible to construct a logically consistent theory of probability. We shall begin the next chapter with the axioms proposed by the Russian mathematician A. N. Kolmogorov.

Examples II

1. Explain what you mean by
   (*i*) Random experiment. [*C.H.* (*Math*) '80, '81, '82, '84, '86, '88]
   [*C.H.* (*Math*) '80]
   (*ii*) Event.
   (*iii*) Event space. [*C.H.* (*Math*) '80, '83, '86]
   (*iv*) Statistical regularity. [*C.H.* (*Math*) '82, '84, '86, '88]
   (*v*) Elementary event. [*C.H.* (*Math*) '83 ; *C.H.* (*Econ*) '81]
   (*vi*) Compound event. [*C.H.* (*Math*) '88 ; *C.H.* (*Econ*) '81]
   (*vii*) Mutually exclusive events. [*C.H.* (*Econ*) '86, '88]
   (*viii*) Exhaustive set of events. [*C.H.* (*Econ*) '86]

2. Give the classical definition of probability of an event and criticize the main drawbacks of the classical theory of probability. [*C.H.* (*Math*) '82 ; *C.H.* (*Econ*) '83, '90]

3. Give the frequency definition of probability of an event and deduce the classical definition from it. Criticize the main drawbacks of the frequency definition of probability. [*C.H.* (*Econ*) '90]

4. Let A, B, C be three arbitrary events. Find expressions for the following events using the usual set theoretic notations :
   (*i*) only A occurs ; (*ii*) both A and B but not C occur ;
   (*iii*) all these events occur ; (*iv*) at least one event occurs ;
   (*v*) at least two events occur. [*C.H.* (*Econ*) '81]

5. Let A, B, C be the events that a man walking on the street will see a new immigrant, a hippie, a tourist from France respectively. Interpret the following events :
   (*i*) $A\cap(B\cap C)$; (*ii*) $A\cap(\overline{B\cup C})$; (*iii*) $(A\cup B)\cap\overline{C}$;
   (*iv*) $A\cup B\cup C$. [*C.H.* (*Econ*) '85]

6. Set up a sample space (*i.e., event space*) for the single toss of a pair of dice. [*C. H.* (*Econ*) '87]

7. If A, B are two events of a random experiment, then express the events (*i*) 'exactly one of A and B' ; (*ii*) 'not more than one of the events A or B occurs' in set theoretic notation.

8. A, B and C are three arbitrary events. Find expressions for the following events, in set theoretic notations :
   (*i*) None occurs, (*ii*) Two and no more occur.

Answers

4. (*i*) $A\overline{B}\,\overline{C}$; (*ii*) $AB\overline{C}$; (*iii*) ABC ; (*iv*) $A+B+C$;
   (*v*) $AB\overline{C}+A\overline{B}C+\overline{A}BC+ABC$.

5. (*i*) will see a new immigrant, a hippie and a tourist from France.
   (*ii*) a new immigrant and no hippie and no tourist.
   (*iii*) a new immigrant or a hippie and no tourist.
   (*iv*) either a new immigrant or a hippie or a tourist from France.

6. See § 2.3.

7. (*i*) $A\overline{B}+\overline{A}B$; (*ii*) $A\overline{B}+\overline{A}B+\overline{A}\,\overline{B}$.

8. (*i*) $\overline{A}\,\overline{B}\,\overline{C}$, (*ii*) $AB\overline{C}+\overline{A}BC+A\overline{B}C$.

CHAPTER III

AN AXIOMATIC CONSTRUCTION OF THE THEORY OF PROBABILITY

3.1. Axiomatic Definition of Probability.

Kolmogorov (1933), in his axiomatic construction of the probability theory, starts from a set S of simple events. The elements of this set are immaterial for the logical development of the theory of probability. Then a class Δ of subsets of S is considered satisfying properties (*i*), (*ii*), (*iii*), given in § 2.4, Chapter II. Elements of the family Δ are called events and Δ will be called a class of events. In the axiomatic theory, since the 'events' do not refer to definite concrete objects, it is not necessary to mention 'random experiment' in formulating the axioms. But we shall always mean 'events' as events of the real world. So we shall state the axioms with reference to a random experiment. Now we can formulate the axioms that define probability :

Let E be a given random experiment and S be the corresponding event space. Also let Δ be the class of subsets of S forming the class of events of E. A mapping $P : \Delta \to R$ is called a probability function defined on Δ and the unique real number $P(A)$ determined by P is called the probability of the event A where $A \in \Delta$ if the following axioms, known as axioms of probability, are satisfied :

Axiom (*a*). $P(A) \geqslant 0$ for every event $A \in \Delta$.

Axiom (*b*). $P(S) = 1$.

Axiom (*c*). If $A_1, A_2, \ldots\ldots, A_n, \ldots$ be countably infinite number of pairwise mutually exclusive events, *i.e.*, if $A_i A_j = O$ whenever $i \neq j$ and $A_i, A_j \in \Delta$,

$$\text{then } P(A_1 + A_2 + A_3 + \cdots + A_n + \cdots) = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots \tag{3.1.1}$$

The entire mathematical theory of probability will be built by three objects, namely (*i*) the event space (*ii*) the class of events Δ (*iii*) the probability function $P : \Delta \to R$. The ordered 3-*tuple* (S, Δ, P) is called a probability space.

It is important to realize that the axioms of probability will not give unique assignment of probabilities to events, *i.e.*, for the same event space S we can choose probabilities in many ways satisfying the axioms. We take an example to illustrate this fact. In the random experiment of throwing a die, the event space S consists of six simple events : $E_1, E_2, E_3, E_4, E_5, E_6$, where E_i signifies the event 'the die shows i points' $(i = 1, 2, 3, 4, 5, 6)$. Now we can assign probabilities to these simple events either

$$P(E_1) = P(E_2) = \cdots\cdots = P(E_6) = \tfrac{1}{6}$$

$$\text{or, } P(E_1) = P(E_2) = P(E_3) = \tfrac{1}{4},\ P(E_4) = P(E_5) = P(E_6) = \tfrac{1}{12},$$

thus satisfying all the axioms of probability in either case. But this apparent incompleteness of a system of axioms in probability does not put any hindrance to our approach towards a consistent and logical theory. In fact, the axioms simply clarify relationships between probabilities that we assign so that we will be consistent with our intuitive notion of probability. Thus if the real number p is assigned to an event A as the probability of A, then the axioms will imply $0 \leq p \leq 1$ and $P(\bar{A}) = 1 - p$, but we do not get any relation between the given random experiment E and the probability p assigned to A and so we do not get any practical meaning of p from the axioms of probability.

So, before going into details of such a theory, we first include the following concept of frequency interpretation of probability, which will connect any probability number defined by the above set of axioms with experimentally measured value of probability.

3.2. Frequency Interpretation of Probability.

To give a practical meaning of $P(A)$ we make the following assumption known as frequency interpretation of probability :

Let the corresponding random experiment E be repeated under identical conditions N times and let the event A occur $N(A)$ times in these N trials of E. Then the frequency ratio $f(A) = \dfrac{N(A)}{N}$ of the event A is approximately equal to $P(A)$ if N is very large.

Remark : Note the difference between frequency interpretation of probability with the frequency definition of probability as discussed in Chapter II. (P-41)

Frequency Interpretation of the Axioms of Probability.

All the axioms $(a)-(c)$ are consistent with the properties of frequency ratio. If the event A occurs $N(A)$ times when the corresponding random expriment E is repeated N times under identical conditions, then the frequency ratio $f(A)=\frac{N(A)}{N}\geq 0$.

Hence, the probability $P(A)$ of the event A being approximately equal to $f(A)$ if N is very large, it follows that $P(A)\geq 0$.

Further, if A is a certain event, $N(A)=N$ and therefore, $f(A)=\frac{N(A)}{N}=1$.

Finally, if the events A_i $(i=1, 2, \cdots\cdots)$ are pairwise mutually exclusive having frequency ratio $f(A_i)$, then the frequency ratio of the event 'at least one of $A_1, A_2, \cdots, A_n, \cdots$ occurs' is

$$f(A_1+A_2+\cdots+A_n+\cdots)=\frac{N(A_1)+N(A_2)+\cdots+N(A_n)+\cdots}{N},$$

where the event A_i occurs $N(A_i)$ times, when the random experiment E is repeated N times under identical conditions

$$=\frac{N(A_1)}{N}+\frac{N(A_2)}{N}+\cdots+\frac{N(A_n)}{N}+\cdots$$
$$=f(A_1)+f(A_2)+\cdots+f(A_n)+\cdots.$$

3.3. Deductions from Axiomatic Definition.

I. $P(O)=0.$ (3.3.1)

We have $O=O+O+O+\cdots$, where O occurs countably infinite number of times in the right hand side.

Here $OO=O$. So by axiom (c),

$$P(O+O+O+\cdots)=P(O)+P(O)+P(O)+\cdots$$

$$\text{or,}\quad P(O)=P(O)+P(O)+\cdots \qquad (3.3.2)$$

Let $P(O)=k$.

Then $k\geq 0$ by axiom (a). If $k\neq 0$, then the infinite series in the right hand side of (3.3.2) is

$k+k+k+\cdots\cdots$ which is divergent,

since here $\underset{n\to\infty}{Lt}\ nk=\infty$ $(\because k>0)$

and then (3.3.2) is impossible. So it is proved that the assumption $k\neq 0$ is wrong. Hence $k=0$. $\therefore$ $P(O)=0$.

We shall later see (P 153) that $P(A)=0$ does not imply that $A=0$.

II. THEOREM 3.3.1. *If $A_1, A_2, \ldots\ldots, A_n$ be finite number of pairwise mutually exclusive events, then*

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n). \qquad (3.3.3)$$

Proof: Let $A_j=O$ for $j=n+1, n+2, \ldots\ldots$

Then applying axiom (c) to countably infinite number of pairwise mutually exclusive events $A_1, A_2, \ldots, A_n, A_{n+1}, \ldots$ we get

$$\begin{aligned}
&P(A_1+A_2+\cdots+A_n)\\
&=P(A_1+A_2+\cdots+A_n+A_{n+1}+\cdots)\\
&=P(A_1)+P(A_2)+\cdots+P(A_n)+P(A_{n+1})+\cdots\\
&=P(A_1)+P(A_2)+\cdots+P(A_n)+P(O)+P(O)+\cdots\\
&=P(A_1)+P(A_2)+\cdots+P(A_n)+0+0+\cdots\\
&=P(A_1)+P(A_2)+\cdots+P(A_n).
\end{aligned}$$

III. $P(\bar{A})=1-P(A).$ (3.3.4)

Proof: We have $A+\bar{A}=S$, where S is the corresponding event space.

$\therefore$ $P(A+\bar{A})=P(S)=1$, by axiom (b).

Now A and $\bar{A}$ are mutually exclusive and hence by (3.3.3), $P(A)+P(\bar{A})=1$.

$\therefore$ $P(\bar{A})=1-P(A)$.

IV. Deduction of the Classical Definition.

In this case the event space S is finite and contains n distinct elements $u_1, u_2, \ldots, u_n$ (say), so that

$$S=\{u_1, u_2, \ldots, u_n\}.$$

Here the n distinct simple events $U_1=\{u_1\}$, $U_2=\{u_2\}$, $\ldots\ldots$, $U_n=\{u_n\}$ are equally likely, which means that the simple events have equal probability, *i.e.*,

$$P(U_1)=P(U_2)=\cdots=P(U_n). \qquad (3.3.5)$$

Since any two distinct simple events are necessarily mutually exclusive, by axiom (b) and (3.3.3) we have

$$1 = P(S) = P(U_1 + U_2 + \cdots + U_n) = P(U_1) + P(U_2) + \cdots + P(U_n). \tag{3.3.6}$$

From (3.3.5) and (3.3.6), we get

$$P(U_1) = P(U_2) = \cdots = P(U_n) = \frac{P(U_1) + P(U_2) + \cdots + P(U_n)}{n} = \frac{1}{n}$$

Let now A be an event connected to the given random experiment. If the event A contains m distinct elements of S, say $u_{i_1}, u_{i_2}, \ldots\ldots u_{i_m}$ where $i_1, i_2, \ldots\ldots, i_m$ take distinct values from the set $\{1, 2, \ldots\ldots, n\}$, we can write

$$A = U_{i_1} + U_{i_2} + \cdots + U_{i_m}.$$

Since any two distinct simple events are mutually exclusive, we get by (3.3.3)

$$P(A) = P(U_{i_1}) + P(U_{i_2}) + \cdots + P(U_{i_m})$$

$$= \frac{1}{n} + \frac{1}{n} + \ldots\ldots m \text{ times}$$

$$= \frac{m}{n}.$$

Hence the classical definition is established.

V. $0 \leq P(A) \leq 1$, for any event A. (3.3.7)

By axiom (*a*), $P(A) \geq 0$, $P(\bar{A}) \geq 0$.

Also by (3.3.4), $1 - P(A) = P(\bar{A}) \geq 0$.

$\therefore \quad P(A) \leq 1.$

Hence $0 \leq P(A) \leq 1$.

VI. If A be a subevent of B, *i.e.*, $A \subseteq B$, then $P(A) \leq P(B)$. (3.3.8)

Since $A \subseteq B$, $B = A + (B - A)$.

Since A and $B - A$ are mutually exclusive, by (3.3.3)

$$P(B) = P(A) + P(B - A).$$

$$\therefore \quad P(B - A) = P(B) - P(A).$$

Since by axiom (*a*), $P(B - A) \geq 0$, it follows that

$$P(A) \leq P(B).$$

Remark : The converse of the above proposition (3.3.8) is not true. This is clear from the following example.

Let E be the random experiment of throwing a die. Let A be the event 'multiple of 3' and B be the event 'even face'.

Then $A = \{3, 6\}$ and $B = \{2, 4, 6\}$

$\therefore \quad P(A) = \frac{2}{6} = \frac{1}{3}$, $P(B) = \frac{3}{6} = \frac{1}{2}$, if we use classical definition of probability.

Hence $P(A) < P(B)$ but $A \not\subset B$.

VII. Addition Rule :

(*i*) **Theorem 3.3.2 :** *If A and B be any two events connected to a random experiment, then*

$$P(A + B) = P(A) + P(B) - P(AB) \tag{3.3.9}$$

i.e., the probability of at least one of the events A and B to occur is $P(A) + P(B) - P(AB)$.

Proof : The events $A - AB$, $B - AB$ and AB are pairwise mutually exclusive events.

Also $A + B = (A - AB) + (B - AB) + AB$.

Hence by (3.3.3),

$$P(A + B) = P\{(A - AB) + (B - AB) + AB\}$$
$$= P(A - AB) + P(B - AB) + P(AB). \tag{3.3.10}$$

Again $A = (A - AB) + AB$, $B = (B - AB) + AB$.

Hence by (3.3.3), as $A - AB$, AB are mutually exclusive and $B - AB$, AB are also mutually exclusive,

$$P(A) = P(A - AB) + P(AB) \tag{3.3.11}$$

$$P(B) = P(B - AB) + P(AB). \tag{3.3.12}$$

Eliminating $P(A - AB)$ and $P(B - AB)$ from (3.3.10), (3.3.11) and (3.3.12) we get

$$P(A + B) = \{P(A) - P(AB)\} + \{P(B) - P(AB)\} + P(AB)$$
$$= P(A) + P(B) - P(AB).$$

(*ii*) **For any three events A, B, C,**

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC) \tag{3.3.13}$$

$$P(A + B + C) = P[(A + B) + C]$$
$$= P(A + B) + P(C) - P[(A + B)C] \text{ by } (3.3.9)$$
$$= P(A) + P(B) - P(AB) + P(C) - P(AC + BC)$$

or, $P(A+B+C)=P(A)+P(B)+P(C)-P(AB)-\{P(AC)+P(BC)-P(ABC)\}$, by (3.3.9)

$$=P(A)+P(B)+P(C)-P(AB)-P(BC)-P(CA)+P(ABC).$$

(iii) General Addition Rule.

THEOREM 3.3.3: *If $A_1, A_2, \ldots, A_n$ be n events (n is a positive integer ≥ 2) connected to a given random experiment E, then*

$$P(A_1+A_2+\cdots+A_n)=\sum_{i_1=1}^{n} P(A_{i_1})-\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{n} P(A_{i_1}A_{i_2})$$

$$+\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{n} P(A_{i_1}A_{i_2}A_{i_3})-\cdots+(-1)^{n-1}P(A_1A_2\ldots A_n). \quad (3.3.14)$$

Proof: For any two events A_1, A_2 we have already proved that $P(A_1+A_2)=P(A_1)+P(A_2)-P(A_1A_2)$, which shows that the proposition (3.3.14) is true for $n=2$.

Let the proposition (3.3.14) be true for $n=m$, where m is a positive integer ≥ 2. Then we have

$$P(A_1+A_2+\cdots+A_m)$$

$$=\sum_{i_1=1}^{m} P(A_{i_1})-\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m} P(A_{i_1}A_{i_2})+\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{m} P(A_{i_1}A_{i_2}A_{i_3})-\cdots$$

$$+(-1)^{m-1}P(A_1A_2\ldots\ldots A_m). \quad (3.3.15)$$

We now consider any $(m+1)$ events $A_1, A_2, \cdots, A_m, A_{m+1}$.

Then $P(A_1+A_2+\cdots+A_m+A_{m+1})$

$$=P[(A_1+A_2+\cdots+A_m)+A_{m+1}]$$

$$=P(A_1+A_2+\cdots+A_m)+P(A_{m+1})-P[(A_1+A_2+\cdots+A_m)A_{m+1}]. \quad (3.3.16)$$

Now $P[(A_1+A_2+\cdots+A_m)A_{m+1}]$

$$=P(A_1A_{m+1}+A_2A_{m+1}+\cdots+A_mA_{m+1})$$

$$=\sum_{i_1=1}^{m} P(A_{i_1}A_{m+1})-\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m} P(A_{i_1}A_{i_2}A_{m+1})$$

$$+\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{m} P(A_{i_1}A_{i_2}A_{i_3}A_{m+1})-\cdots\cdots$$

$$\cdots\cdots+(-1)^{m-1}P(A_1A_2\cdots A_mA_{m+1}) \quad (3.3.17)$$

since the proposition is assumed to be true for any m events.

Then from (3.3.15), (3.3.16) and (3.3.17) we get

$$P(A_1+A_2+\cdots\cdots+A_m+A_{m+1})$$

$$=\sum_{i_1=1}^{m} P(A_{i_1})-\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m} P(A_{i_1}A_{i_2})+\cdots+(-1)^{m-1}P(A_1A_2\cdots A_m)$$

$$+P(A_{m+1})-\Big[\sum_{i_1=1}^{m} P(A_{i_1}A_{m+1})-\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m} P(A_{i_1}A_{i_2}A_{m+1})$$

$$+\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{m} P(A_{i_1}A_{i_2}A_{i_3}A_{m+1})-\cdots\cdots$$

$$+(-1)^{m-1}P(A_1A_2\cdots\cdots A_mA_{m+1})\Big]$$

$$=\Big[\sum_{i_1=1}^{m} P(A_{i_1})+P(A_{m+1})\Big]$$

$$-\Big[\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m} P(A_{i_1}A_{i_2})+\sum_{i_1=1}^{m} P(A_{i_1}A_{m+1})\Big]$$

$$+\Big[\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{m} P(A_{i_1}A_{i_2}A_{i_3})+\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m} P(A_{i_1}A_{i_2}A_{m+1})\Big]$$

$$-\Big[\sum_{\substack{i_1,i_2,i_3,i_4=1\\(i_1<i_2<i_3<i_4)}}^{m} P(A_{i_1}A_{i_2}A_{i_3}A_{i_4})$$

$$+\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{m} P(A_{i_1}A_{i_2}A_{i_3}A_{m+1})\Big]+\cdots\cdots$$

$$+\Big[(-1)^{m-1}P(A_1A_2\cdots A_m)$$

$$-(-1)^{m-2}\sum_{\substack{i_1,i_2,\ldots,i_{m-1}=1\\(i_1<i_2<\ldots<i_{m-1})}}^{m} P(A_{i_1}A_{i_2}\cdots A_{i_{m-1}}A_{m+1})\Big]$$

$$-(-1)^{m-1}P(A_1A_2\cdots A_mA_{m+1})$$

$$=\sum_{i_1=1}^{m+1}P(A_{i_1})-\sum_{\substack{i_1,i_2=1\\(i_1<i_2)}}^{m+1}P(A_{i_1}A_{i_2})+\sum_{\substack{i_1,i_2,i_3=1\\(i_1<i_2<i_3)}}^{m+1}P(A_{i_1}A_{i_2}A_{i_3})$$

$$-\cdots\cdots+(-)^{m-1}\sum_{\substack{i_1,i_2,\ldots,i_m=1\\(i_1<i_2<\cdots<i_m)}}^{m+1}P(A_{i_1}A_{i_2}\cdots A_{i_m})$$

$$+(-1)^mP(A_1A_2\cdots A_{m+1}).$$

Hence the proposition (3.3.14) is true for $n=m+1$ if it is true for $n=m$ where $m\ (\geq 2)$ is a positive integer. Also the proposition is true for $n=2$. Hence by induction principle, the proposition (3.3.14) is true for any integer $n\geq 2$. So the theorem is proved.

3.4. Some Important Inequalities.

I. Boole's Inequality :

Theorem 3.4.1. *If $A_1, A_2, \ldots\ldots A_n$ be any n events connected to a random experiment E, then*

$$P(A_1+A_2+\cdots\cdots+A_n)\leq P(A_1)+P(A_2)+\cdots\cdots+P(A_n). \qquad (3.4.1)$$

Proof : We have for any two events A_1, A_2,

$$P(A_1+A_2)=P(A_1)+P(A_2)-P(A_1A_2)$$
$$\leq P(A_1)+P(A_2),$$

since by axiom (*a*), $P(A_1A_2)\geq 0$.

$\therefore$ the given proposition is true for $n=2$.

We now assume that the inequality is true for $n=m$, where m is a positive integer ≥ 2.

We now consider any $(m+1)$ events $A_1, A_2, \ldots\ldots A_{m+1}$. By hypothesis,

$$P(A_1+A_2+\cdots+A_m)\leq P(A_1)+P(A_2)+\cdots+P(A_m). \qquad (3.4.2)$$

Now $P(A_1+A_2+\cdots+A_m+A_{m+1})$

$=P[(A_1+A_2+\cdots+A_m)+A_{m+1}]$

$\leq P(A_1+A_2+\cdots+A_m)+P(A_{m+1})$,

since the inequality is true for $n=2$

$\leq P(A_1)+P(A_2)+\cdots+P(A_m)+P(A_{m+1})$, by (3.4.2).

Thus the inequality is true for $n=m+1$ whenever it is true for $n=m$, where m is a positive integer ≥ 2. We have already shown that the inequality is true for $n=2$. Hence by the principle of mathematical induction, we conclude that the given inequality is true for all $n\geq 2$, n being a positive integer. We note that (3.4.1) is true for $n=1$ with the sign of equality.

II. Bonferroni's Inequalities.

Theorem 3.4.2 : *If $A_1, A_2, \ldots, A_n$ be any n events connected to a random experiment E, then*

$$(i)\quad P(A_1A_2\ldots\ldots A_n)\geq 1-\sum_{i=1}^{n}P(\bar{A}_i) \qquad (3.4.3)$$

$$(ii)\quad P(A_1A_2\ldots\ldots A_n)\geq \sum_{i=1}^{n}P(A_i)-(n-1). \qquad (3.4.4)$$

Proof : (*i*) We have by Boole's inequality

$$P(\bar{A}_1+\bar{A}_2+\cdots+\bar{A}_n)\leq P(\bar{A}_1)+P(\bar{A}_2)+\cdots+P(\bar{A}_n)$$

or. $P(\overline{A_1A_2\ldots\ldots A_n})\leq P(\bar{A}_1)+P(\bar{A}_2)+\cdots+P(\bar{A}_n)$

by De Morgan's law.

or, $1-P(A_1A_2\ldots\ldots A_n) \le P(\bar{A}_1)+P(\bar{A}_2)+\cdots+P(\bar{A}_n)$

or, $P(A_1A_2\ldots\ldots A_n) \geq 1-\sum_{i=1}^{n} P(\bar{A}_i)$

which proves (*i*).

(*ii*) Now $P(\bar{A}_i)=1-P(A_i)$.

$$\therefore \quad 1-\sum_{i=1}^{n} P(\bar{A}_i)=1-\sum_{i=1}^{n}\{1-P(A_i)\}$$

$$=\sum_{i=1}^{n} P(A_i)-(n-1).$$

Hence the result (*ii*) follows.

3.5. Limit of a Sequence of Events.

Let $\{A_n\}$ be a sequence of events connected to a given random experiment E. Then $\prod_{n=1}^{\infty}\left(\sum_{k=n}^{\infty} A_k\right)$ is an event connected to E and it is called the superior limit of $\{A_n\}$ and it is denoted by $\overline{\lim\limits_{n\to\infty}}\,(A_n)$.

Also $\sum_{n=1}^{\infty}\left(\prod_{k=n}^{\infty} A_k\right)$ is an event connected to E and it is called the inferior limit of $\{A_n\}$ and is denoted $\underline{\lim\limits_{n\to\infty}}\,(A_n)$.

A sequence of events $\{A_n\}$ is said to be convergent if and only if

$$\underline{\lim_{n\to\infty}}\,(A_n)=\overline{\lim_{n\to\infty}}\,(A_n) \tag{3.5.1}$$

and in this case $\overline{\lim\limits_{n\to\infty}}\,(A_n)$ or $\underline{\lim\limits_{n\to\infty}}\,(A_n)$ is called the limit of the sequence $\{A_n\}$ and it is denoted by $\lim\limits_{n\to\infty} A_n$.

If $\{A_n\}$ be a monotonically increasing sequence, *i.e.*, $A_n \subseteq A_{n+1}$ for all n, then it can be shown that

$$\overline{\lim_{n\to\infty}}\, A_n=\underline{\lim_{n\to\infty}}\, A_n=\sum_{n=1}^{\infty} A_n=\lim_{n\to\infty}(A_n). \tag{3.5.2}$$

If $\{A_n\}$ be a monotonically decreasing sequence, *i.e.*, $A_{n+1} \subseteq A_n$ for all n, then it can be shown that

$$\overline{\lim_{n\to\infty}}\, A_n=\underline{\lim_{n\to\infty}}\, A_n=\prod_{n=1}^{\infty} A_n=\lim_{n\to\infty} A_n. \tag{3.5.3}$$

3.6. THEOREM 3.6.1. *If $\{A_n\}$ be a monotone sequence of events, then*

$$P(\lim_{n\to\infty} A_n)=\lim_{n\to\infty} P(A_n). \tag{3.6.1}$$

Proof: Case 1. Let $\{A_n\}$ be a monotonically increasing sequence of events, *i.e.*, $A_n \subseteq A_{n+1}$ for all n. We define another sequence of events $\{B_n\}$ as follows:

$B_1=A_1$, $B_2=A_2-A_1$, $B_3=A_3-A_2$, ..., $B_n=A_n-A_{n-1}$ and so on, when $n \geq 2$. We first show that B_i's are mutually exclusive. If possible, let $B_iB_j \neq O$, where $i \neq j$. Then there exists an element ω such that $\omega \in B_i$, $\omega \in B_j$. Without any loss of generality we assume that $i > j$. Then $i=j+k$, where k is a positive integer.

Now $\omega \in B_i$ implies that $\omega \in A_i - A_{i-1}$, *i.e.*, $\omega \in A_i$ and $\omega \notin A_{i-1}$. Again $j=i-k \leq i-1$.

Since $\{A_n\}$ is monotonically increasing, $i-1 \geq j$ implies that $A_j \subseteq A_{i-1}$. Now $\omega \in B_j=A_j-A_{j-1}$ implies $\omega \in A_j \subseteq A_{i-1}$ if $j>1$ and $\omega \in B_j=A_j \subseteq A_{i-1}$ if $j=1$.

$\therefore$ $\omega \in A_{i-1}$ which is a contradiction.

$\therefore$ $B_iB_j=O$ whenever $i \neq j$.

We now show that $\sum_{n=1}^{\infty} A_n=\sum_{n=1}^{\infty} B_n$.

Let $a \in \sum_{n=1}^{\infty} A_n$. Then there exists a positive integer m such that $a \in A_m$. Now consider the set

$Q=\{n : n \text{ is a positive integer and } a \in A_n\}$.

Evidently $m \in Q$ and so Q is a non-empty set of positive integers. Then by well-ordering principle of natural numbers, Q has a least element p (say). Then if $p=1$, $a \in A_1$ and if $p>1$, then $a \in A_p$, but $a \notin A_{p-1}$.

Now if $p=1$, $a \in A_1=B_1$ and if $p>1$, then $a \in A_p$ but $a \notin A_{p-1}$ and so $a \in A_p-A_{p-1}=B_p$. In either case $a \in \sum_{n=1}^{\infty} B_n$.

Thus $a \in \sum_{n=1}^{\infty} A_n$ implies $a \in \sum_{n=1}^{\infty} B_n$. $\therefore \sum_{n=1}^{\infty} A_n \subseteq \sum_{n=1}^{\infty} B_n$. (3.6.2)

Again let $b \in \sum_{n=1}^{\infty} B_n$. Then $b \in B_k$ for some positive integer k. If $k=1$, then $b \in B_1 = A_1$. If $k > 1$, $b \in B_k = A_k - A_{k-1}$ and so $b \in A_k$. In either case $b \in \sum_{n=1}^{\infty} A_n$.

Thus $b \in \sum_{n=1}^{\infty} B_n$ implies $b \in \sum_{n=1}^{\infty} A_n$. $\therefore \sum_{n=1}^{\infty} B_n \subseteq \sum_{n=1}^{\infty} A_n$. (3.6.3)

From (3.6.2) and (3.6.3) we conclude that

$$\sum_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} B_n. \quad (3.6.4)$$

It can be similarly shown that

$$\sum_{i=1}^{n} A_i = \sum_{j=1}^{n} B_j \text{ for all } n. \quad (3.6.5)$$

Again since $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots\cdots \subseteq A_n$.

$$\sum_{i=1}^{n} A_i = A_n. \quad (3.6.6)$$

Now $P(\lim_{n\to\infty} A_n) = P(\sum_{n=1}^{\infty} A_n)$, by (3.5.2)

$= P(\sum_{n=1}^{\infty} B_n)$, by (3.6.4)

$= \sum_{n=1}^{\infty} P(B_n)$,

$B_1, B_2, \ldots \ldots B_n, \ldots \ldots$, being pairwise mutually exclusive, we apply axiom (*c*).

$= \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k)$

$= \lim_{n\to\infty} [P(B_1)+P(B_2)+\cdots+P(B_n)]$

$= \lim_{n\to\infty} [P(B_1+B_2+\cdots+B_n)]$, by (3.3.3)

or, $P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} [P(A_1+A_2+\cdots+A_n)]$, by (3.6.5)

$= \lim_{n\to\infty} P(A_n)$, by (3.6.6)

$\therefore P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$.

Case II. Let $\{A_n\}$ be a monotonically decreasing sequence of events. Then $A_n \supseteq A_{n+1}$ for all n.

$\therefore \bar{A}_n \subseteq \bar{A}_{n+1}$ for all n.

i.e., $\{\bar{A}_n\}$ is monotonically increasing.

Then by Case I, $P(\lim_{n\to\infty} \bar{A}_n) = \lim_{n\to\infty} P(\bar{A}_n)$

or, $P\left(\sum_{n=1}^{\infty} \bar{A}_n\right) = \lim_{n\to\infty} [1-P(A_n)]$, by (3.5.2)

or, $P\left(\overline{\prod_{n=1}^{\infty} A_n}\right) = 1 - \lim_{n\to\infty} P(A_n)$,

by De Morgan's law

or, $1 - P\left(\prod_{n=1}^{\infty} A_n\right) = 1 - \lim_{n\to\infty} P(A_n)$

$\therefore P\left(\prod_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$.

But here $\lim_{n\to\infty} A_n = \prod_{n=1}^{\infty} A_n$, by (3.5.3)

$\therefore P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$.

This completes the proof of the theorem.

3.7. Conditional Probability.

The concept of conditional probability is already introduced in § 2.12 on the basis of frequency definition of probability and the expression obtained for it suggests the following definition of conditional probability in the axiomatic theory.

Let A and B be any two events connected to a given random experiment E. The conditional probability of the event A on the hypothesis that the event B has occurred, denoted by $P(A \mid B)$, is defined as

$$P(A \mid B) = \frac{P(AB)}{P(B)} \quad (3.7.1)$$

provided $P(B) \neq 0$.

Similarly we can define $P(B \mid A)$ when $P(A) \neq 0$.

For example, let us consider the random experiment of throwing a symmetric die. If A and B be the events 'even face' and 'multiple of three', let us find $P(A \mid B)$ and $P(B \mid A)$. Here the event space contains 6 simple events and the number of simple events favourable to the events A, B and AB are respectively 3, 2 and 1.

$$\therefore \quad P(A)=\tfrac{1}{2},\ P(B)=\tfrac{1}{3},\ P(AB)=\tfrac{1}{6}.$$

So by definition $P(B \mid A)=\dfrac{P(AB)}{P(A)}=\dfrac{1}{3}$

and $$P(A \mid B)=\frac{P(AB)}{P(B)}=\frac{1}{2}.$$

3.8. Frequency Interpretation.

For two events A and B connected to a random experiment E, let the events B and AB occur $N(B)$ and $N(AB)$ times respectively, when the random experiment E is repeated N times under identical conditions. Then the conditional frequency ratio

$$f(A \mid B)=\frac{N(AB)}{N(B)} \tag{3.8.1}$$

is approximately equal to the conditional probability $P(A \mid B)$ for a long sequence of repetitions of the random experiment E under identical conditions, *i.e.*, when N is very large, provided $P(B) \neq 0$.

3·9. THEOREM 3.9.1. *The conditional probability satisfies all the axioms of probability.*

Proof: (*i*) We see that for any two events A, B with $P(B) \neq 0$,

$$P(A \mid B)=\frac{P(AB)}{P(B)}.$$

Now by axiom (*a*) $P(AB) \geq 0$, $P(B) > 0$.

Hence, $P(A \mid B) \geq 0$.

(*ii*) Let S be the event space. If $P(B) \neq 0$, then

$$P(S \mid B)=\frac{P(SB)}{P(B)}=\frac{P(B)}{P(B)}=1.$$

(*iii*) Let $A_1, A_2, \ldots\ldots, A_n, \ldots\ldots$ be countably infinite number of pairwise mutually exclusive events, connected to the given random experiment. Now we have for any event B, with $P(B) \neq 0$,

$$P\{(A_1+A_2+\cdots\cdots+A_n+\cdots\cdots) \mid B\}$$
$$=\frac{P\{(A_1+A_2+\cdots\cdots+A_n+\cdots\cdots)B\}}{P(B)}$$
$$=\frac{P(A_1B+A_2B+\cdots\cdots+A_nB+\cdots\cdots)}{P(B)}$$

Since the events $A_1, A_2, \ldots\ldots, A_n, \ldots$ are mutually exclusive, $A_1B, A_2B, \ldots\ldots, A_nB, \ldots$ are also pairwise mutually exclusive events $[(A_iB)(A_jB)=(A_iA_j)B=OB=O,\ i \neq j]$.

$$\therefore \quad P(A_1B+A_2B+\cdots\cdots+A_nB+\cdots\cdots)$$
$$=P(A_1B)+P(A_2B)+\cdots\cdots+P(A_nB)+\cdots \quad \text{[by axiom (c).]}$$

Hence $P\{(A_1+A_2+\cdots\cdots+A_n+\cdots\cdots) \mid B\}$

$$=\frac{P(A_1B)+P(A_2B)+\cdots\cdots+P(A_nB)+\cdots\cdots}{P(B)}$$
$$=\frac{P(A_1B)}{P(B)}+\frac{P(A_2B)}{P(B)}+\cdots\cdots+\frac{P(A_nB)}{P(B)}+\cdots\cdots$$
$$=P(A_1 \mid B)+P(A_2 \mid B)+\cdots\cdots+P(A_n \mid B)+\cdots\cdots$$

Hence the conditional probability statisfies all the axioms of probability.

3.10. THEOREM 3.10.1. *If $A_1, A_2, \ldots\ldots\ldots, A_n$ be pairwise mutually exclusive events, one of which certainly occurs (i.e., $A_1, A_2, \ldots\ldots, A_n$ form an exhaustive set of events), then*

$$P(B)=\sum_{i=1}^{n} P(A_i)P(B \mid A_i), \tag{3.10.1}$$

where B is any event connected to the same random experiment, provided the conditional probabilities are defined.

Proof: Here $A_1, A_2, \ldots\ldots, A_n$ are pairwise mutually exclusive events, connected to a random experiment E, *i.e.*, $A_iA_j=O$, $i \neq j$; $i, j=1, 2, \ldots\ldots n$.

Also $S=A_1+A_2+\cdots\cdots+A_n$, since one of $A_1, A_2, \ldots\ldots, A_n$ certainly occurs. We have

$$B=SB=(A_1+A_2+\cdots+A_n)\,B=A_1B+A_2B+\cdots+A_nB.$$

Since $(A_iB)(A_jB)=(A_iA_j)\,B=OB=O$. $(i \neq j;\ i,\ j=1, 2, \ldots\ldots n)$, $A_1B, A_2B, \ldots\ldots, A_n B$ are pairwise mutually exclusive events and hence

$$P(B)=P(A_1B+A_2B+\cdots\cdots+A_nB)$$
$$=P(A_1B)+P(A_2B)+\cdots\cdots P(A_nB).$$

Since $P(A_iB)=P(A_i)P(B \mid A_i)$ for $i=1, 2, \ldots\ldots n$, we get

$$P(B)=\sum_{i=1}^{n} P(A_iB)$$

$$=\sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$

Hence the theorem.

Theorem 3.10.2. Bayes' Theorem.

Let $A_1, A_2, \ldots\ldots, A_n$ be n pairwise mutually exclusive events connected to a random experiment E where at least one of $A_1, A_2, \ldots, A_n$ is sure to happen (i.e., $A_1, A_2, \ldots\ldots, A_n$ form an exhaustive set of n events). Let X be an arbitrary event connected to E, where $P(X) \neq 0$. Also let the probabilities $P(X \mid A_1), P(X \mid A_2), \ldots\ldots, P(X \mid A_n)$ be all known.

Then $$P(A_i \mid X)=\frac{P(A_i)P(X \mid A_i)}{\sum_{r=1}^{n} P(A_r)P(X \mid A_r)}, \quad i=1, 2, \ldots\ldots, n. \tag{3.10.2}$$

Proof: $A_1, A_2, \ldots\ldots, A_n$ being an exhaustive set of events,

$$S=A_1+A_2+\cdots\cdots+A_n$$

where S is the corresponding event space.

$\therefore \; X(A_1+A_2+\cdots\cdots+A_n)=XS=X \quad \because \; X \subseteq S.$

or, $\quad XA_1+XA_2+\cdots\cdots+XA_n=X.$

Now $\quad (XA_i)(XA_j)=X(A_iA_j)=X0=0$ for $i \neq j$,

since $\quad A_iA_j=0$ for $i \neq j$.

$\therefore \; XA_1, XA_2, \ldots\ldots, XA_n$ are pairwise mutually exclusive events and hence

$$P(XA_1)+P(XA_2)+\cdots\cdots+P(XA_n)=P(X)$$

or, $$P(A_1)P(X \mid A_1)+P(A_2)P(X \mid A_2)+\cdots\cdots+P(A_n)P(X \mid A_n)=P(X). \tag{3.10.3}$$

$$\therefore \; P(A_i \mid X)=\frac{P(A_iX)}{P(X)}, \; \because \; P(X) \neq 0$$

$$=\frac{P(A_i)P(X \mid A_i)}{\sum_{r=1}^{n} P(A_r)P(X \mid A_r)}, \text{ for } i=1, 2, \ldots\ldots, n$$

by (3.10.3).

Hence the theorem.

Note 1. Using (3.10.1) we can find $P(B)$ when the conditional probabilities $P(B \mid A_i)(i=1, 2, \ldots\ldots, n)$ can be conveniently obtained.

Note 2. The formula (3.10.1) and (3.10.2) may be extended to an infinite sequence of events $\{A_n\}$.

Note 3. If Y be any event, we have also

$$P(Y \mid X)=\sum_{i=1}^{n} P(YA_i \mid X)=\sum_{i=1}^{n} P(A_i \mid X)P(Y \mid XA_i)$$

$$=\frac{\sum_{i=1}^{n} P(A_i)P(X \mid A_i)P(Y \mid XA_i)}{\sum_{r=1}^{n} P(A_r)P(X \mid A_r)}, \tag{3.10.4}$$

by (3.10.2).

3.11. Independence of Events.

Let A and B be two events connected to a given random experiment. If $P(B) \neq 0$ then $P(A \mid B)$ can be defined and in this case if $P(A \mid B)=P(A)$, then we can say that the probability of A does not depend on the happening of B, *i.e.*, there is one kind of independence between A and B. Also if $P(A) \neq 0$, then $P(B \mid A)$ can be defined and in this case if $P(B \mid A)=P(B)$, we can say that the probability of B does not depend on the happening of A, *i.e.*, there is one kind of independence between A and B. We observe that

$$P(A \mid B)=P(A), \qquad P(B \mid A)=P(B)$$

both lead to $P(AB)=P(A)\,P(B)$.

So formally we can define independence of two events as follows:

Two events A, B are said to be stochastically independent or statistically independent or simply independent if and only if

$$P(AB)=P(A)P(B). \tag{3.11.1}$$

If $P(AB) \neq P(A)P(B)$, then A, B are said to be dependent.

3.12. Mutual and Pairwise Independence of more than two Events.

Three events A, B, C are said to be pairwise independent if

$$\begin{aligned} P(AB) &= P(A)P(B) \\ P(BC) &= P(B)P(C) \\ P(CA) &= P(C)P(A) \end{aligned} \tag{3.12.1}$$

and A, B, C are said to be mutually independent if

$$\begin{aligned} P(AB) &= P(A)P(B) \\ P(BC) &= P(B)P(C) \\ P(CA) &= P(C)P(A) \\ P(ABC) &= P(A)P(B)P(C). \end{aligned} \tag{3.12.2}$$

In general n events $A_1, A_2, \ldots\ldots, A_n$ $(n > 2)$ are said to be mutually independent if

$P(A_iA_j) = P(A_i)P(A_j)$, where $i < j$; i, j any combination of 1, 2, ..., n taken two at a time.

$P(A_iA_jA_k) = P(A_i)P(A_j)P(A_k)$, where $i < j < k$; i, j, k any combination of 1, 2, ..., n taken 3 at a time.

... ...

... ...

$$P(A_1A_2 \ldots A_n) = P(A_1)P(A_2) \ldots P(A_n). \tag{3.12.3}$$

Note 1. From (3.12.3) we see that in defining mutual independence of n events $(n > 2)$,

$${}^nC_2 + {}^nC_3 + \cdots + {}^nC_n = 2^n - n - 1$$

relations are required.

Note 2. From the definition of mutual independence, we see that mutual independence implies pairwise independence, but the converse is not true, as shown by the following example :

Let the equally likely outcomes of an experiment be one of the four points in the three-dimensional space with rectangular co-ordinates (1, 0, 0), (0, 1, 0), (0, 0, 1) and (1, 1, 1). Let A, B, C denote the events 'x-co-ordinate 1', 'y-co-ordinate 1' and 'z-co-ordinate 1' respectively.

Then by using classical definition,

$$\begin{aligned} &P(A) = \tfrac{2}{4} = \tfrac{1}{2}, \quad P(B) = \tfrac{1}{2}, \quad P(C) = \tfrac{1}{2}. \\ &P(AB) = \tfrac{1}{4} = P(A)P(B) \\ &P(BC) = \tfrac{1}{4} = P(B)P(C) \\ &P(CA) = \tfrac{1}{4} = P(C)P(A). \end{aligned}$$

Hence, A, B, C are pairwise independent.

But $P(ABC) = \frac{1}{4}$.

$\therefore$ $P(ABC) \neq P(A)P(B)P(C)$,

which implies that A, B, C are not mutually independent.

Hence, pairwise independence does not always imply mutual independence.

Note 3. It is to be noted that the concept of mutually exclusive events and independent events are not equivalent. We bring out the difference between the two ideas.

If two events A and B are mutually exclusive then $AB = O$ and so the occurrence of one of the two events, in this case, is hindered by anticipating the occurrence of the other.

On the other hand, if the occurrence of one event has no effect on the probability of the occurrence of the other event, the two events are said to be independent and in this case $P(AB) = P(A)P(B)$.

Two events can be mutually exclusive and not independent. For example, consider the random experiment of tossing of two coins. Let A and B be the events 'both the coins show head' and 'both the coins show tail' respectively. Then A and B are clearly mutually exclusive, since if A happens B cannot happen and as such $AB = O$. But $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{4}$. $\therefore$ $P(AB) = 0 \neq P(A)P(B)$, *i.e.*, A and B are not independent.

Again, two events can be independent and not mutually exclusive. For example, consider the random experiment of throwing 2 dice together. Let A and B be the events '6 appears in the first die' and '6 appears in the second die' respectively.

Then $P(AB) = \frac{1}{36} = P(A)P(B) = \frac{1}{6} \times \frac{1}{6}$ and so A and B are independent. Also $AB = \{(6, 6)\} \neq O$ which implies that A and B are not mutually exclusive.

Finally, two events A and B can be both mutually exclusive and independent when

$$P(AB)=P(A)P(B)=0,$$

which holds if at least one of the two events A and B has zero probability. In fact, two events having both non-zero probabilities cannot be simultaneously mutually exclusive and independent.

3.13. General Multiplication Rule.

THEOREM 3.13.1. *If $A_1, A_2, \ldots\ldots, A_n$ $(n \geq 2)$ be n events connected to a random experiment E, then*

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\cdots P(A_n \mid A_1A_2\cdots A_{n-1}) \tag{3.13.1}$$

provided the conditional probabilities are defined.

Proof: For two events A_1, A_2 we have, by the definition of conditional probability,

$$P(A_2 \mid A_1)=\frac{P(A_1A_2)}{P(A_1)}.$$

$$\therefore\ P(A_1A_2)=P(A_1)P(A_2 \mid A_1).$$

Hence the proposition (3.13.1) is true for $n=2$.

Let the proposition be true for $n=m$, where m is a positive integer ≥ 2. Then we have for any m events $A_1, A_2, \ldots A_m$,

$$P(A_1A_2A_3\ldots A_m)=P(A_1)P(A_2 \mid A_1)\ldots P(A_m \mid A_1A_2\cdots A_{m-1}). \tag{3.13.2}$$

Now we consider the $(m+1)$ events $A_1, A_2, \ldots, A_m, A_{m+1}$.

Then $P(A_1A_2\ldots\ldots A_mA_{m+1})$

$=P[(A_1A_2\ldots\ldots A_m)A_{m+1}]$

$=P(A_1A_2\ldots\ldots A_m)P(A_{m+1} \mid A_1A_2\ldots\ldots A_m)$,

since the proposition is true for $n=2$

$=P(A_1)P(A_2 \mid A_1)\ldots\ldots P(A_m \mid A_1A_2\ldots\ldots A_{m-1})$

$P(A_{m+1} \mid A_1A_2\ldots\ldots A_m)$,

by (3.13.2)

This shows that the proposition (3.13.1) is true for $n=m+1$ if it is true for $n=m$. But the proposition is true for $n=2$. Hence by the principle of mathematical induction, the proposition is true for any positive integer $n \geq 2$.

Particular Case.

For three events A_1, A_2, A_3 connected to a random experiment E,

$$\begin{aligned} P(A_1A_2A_3) &= P(BA_3), \text{ where } B=A_1A_2 \\ &= P(B)P(A_3 \mid B) \\ &= P(A_1A_2)P(A_3 \mid A_1A_2) \\ &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2). \end{aligned}$$

3.14. Illustrative Examples.

Ex. 1. *Show that $P(AB) \geq P(A)+P(B)-1$.* [C.H. (Math.) '70]

Since $P(A+B) \leq 1$, $-P(A+B) \geq -1$.

$\therefore$ $P(A+B)=P(A)+P(B)-P(AB)$ gives

$$P(AB)=P(A)+P(B)-P(A+B) \geq P(A)+P(B)-1.$$

Ex. 2. *Show that the probability that exactly one of the events A and B occurs is $P(A)+P(B)-2P(AB)$.* [C.H. (Math.) '83, '86, '90]

We are to find the probability of $A\bar{B}+\bar{A}B$. Since the two events $A\bar{B}$ and $\bar{A}B$ are mutually exclusive,

$$P(\bar{A}B+A\bar{B})=P(\bar{A}B)+P(A\bar{B}).$$

Again $A=A\bar{B}+AB$ and $B=\bar{A}B+AB$. Since $A\bar{B}$, $\bar{A}B$ and AB are pairwise mutually exclusive,

$$P(A)=P(A\bar{B})+P(AB),\ P(B)=P(\bar{A}B)+P(AB)$$

$$\begin{aligned} \therefore\ P(\bar{A}B)+P(A\bar{B}) &= P(A)-P(AB)+P(B)-P(AB) \\ &= P(A)+P(B)-2P(AB) \end{aligned}$$

$$\therefore\ P(\bar{A}B+A\bar{B})=P(A)+P(B)-2P(AB).$$

Ex. 3. *Obtain $P(\bar{A}+B)$, $P(A+\bar{B})$, $P(\bar{A}+\bar{B})$ in terms of $P(A)$, $P(B)$ and $P(AB)$.* [C.H. (Math.) '81]

$$\begin{aligned} P(\bar{A}+B) &= P(\bar{A})+P(B)-P(\bar{A}B) \\ &= 1-P(A)+P(B)-P(\bar{A}B). \end{aligned}$$

Now $\bar{A}B+AB=B$ where $\bar{A}B$ and AB are mutually exclusive.

$\therefore$ $P(\bar{A}B)=P(B)-P(AB)$.

$\therefore$ $P(\bar{A}+B)=1-P(A)+P(B)-P(B)+P(AB)=1-P(A)+P(AB)$.

Again $P(A+\bar{B})=P(A)+P(\bar{B})-P(A\bar{B})$ and as before,

$$P(A\bar{B})+P(AB)=P(A).$$

$\therefore$ $P(A+\bar{B})=P(A)+P(\bar{B})-P(A)+P(AB)=1-P(B)+P(AB)$.

Finally, $P(\bar{A}+\bar{B})=P(\overline{AB})$, by De Morgan's law

$=1-P(AB)$.

Ex. 4. *If A and B are two events such that $P(A)=P(B)=1$, then show that $P(A+B)=1$, $P(AB)=1$.* [C.H. (Math.) '80]

$P(A+B)=P(A)+P(B)-P(AB)=2-P(AB)$.

Now $P(AB)\leq 1$, i.e., $-P(AB)\geq -1$.

$\therefore\ 2-P(AB)\geq 1$ i.e., $P(A+B)\geq 1$.

But $P(A+B)\leq 1$. This implies that $P(A+B)=1$.

Again $2-P(AB)=P(A+B)=1$ $\therefore$ $P(AB)=1$.

Ex. 5. *Establish the inequalities :*

$P(ABC)\leq P(AB)\leq P(A+B)\leq P(A+B+C)\leq P(A)+P(B)+P(C)$ [C.H. (Math.) '79]

for any three events A, B, C.

We have $ABC\subseteq AB$, $\therefore$ $P(ABC)\leq P(AB)$. ... (1)

Again $AB\subseteq A+B$ $\therefore$ $P(AB)\leq P(A+B)$. ... (2)

Also $A+B\subseteq A+B+C$. $\therefore$ $P(A+B)\leq P(A+B+C)$. ... (3)

Finally,

$$P(A+B+C)=P(A+B)+P(C)-P[(A+B)C]$$
$$\leq P(A+B)+P(C),\quad \because P[(A+B)C]\geq 0$$
$$=P(A)+P(B)+P(C)-P(AB)$$
$$\leq P(A)+P(B)+P(C), \quad \ldots (4)$$
$$\because\ P(AB)\geq 0.$$

From (1), (2), (3) and (4), we get

$$P(ABC)\leq P(AB)\leq P(A+B)\leq P(A+B+C)$$
$$\leq P(A)+P(B)+P(C).$$

Ex. 6. *If $P(A\mid C)\geq P(B\mid C)$ and $P(A\mid \bar{C})\geq P(B\mid \bar{C})$, then prove that $P(A)\geq P(B)$.*

From $P(A\mid C)\geq P(B\mid C)$, we get

$$\frac{P(AC)}{P(C)}\geq\frac{P(BC)}{P(C)}$$

$\therefore$ $P(AC)\geq P(BC)$, ... (1)

since $P(C)>0$.

Similarly, from $P(A\mid \bar{C})\geq P(B\mid \bar{C})$ we get

$$\frac{P(A\bar{C})}{P(\bar{C})}\geq\frac{P(B\bar{C})}{P(\bar{C})}.$$

$\therefore$ $P(A\bar{C})\geq P(B\bar{C})$, ... (2)

since $P(\bar{C})>0$.

From (1) and (2) we get $P(AC)+P(A\bar{C})\geq P(BC)+P(B\bar{C})$

or, $P(AC+A\bar{C})\geq P(BC+B\bar{C})$,

since $(AC)(A\bar{C})=0$, $(BC)(B\bar{C})=0$.

$\therefore$ $P(A)\geq P(B)$.

Ex. 7. *If $P(A\mid B)=1$, then prove that $P(ABC)=P(BC)$.*

Here $P(A\mid B)=1$.

$$\therefore\ \frac{P(AB)}{P(B)}=1,\ \text{i.e.,}\ P(AB)=P(B). \quad \ldots (1)$$

Now $ABC+\bar{A}BC=BC$, where ABC and $\bar{A}BC$ are mutually exclusive.

$\therefore$ $P(BC)=P(ABC)+P(\bar{A}BC)$. ... (2)

Again $\bar{A}B+AB=B$ where $\bar{A}B$ and AB are mutually exclusive.

$\therefore$ $P(B)=P(\bar{A}B)+P(AB)$. ... (3)

From (1) and (3), $P(\bar{A}B)=0$. ... (4)

Now $\bar{A}BC\subseteq\bar{A}B$ $\therefore$ $P(\bar{A}BC)\leq P(\bar{A}B)=0$, but $P(\bar{A}BC)\geq 0$.

Hence $P(\bar{A}BC)=0$. ... (5)

From (2) and (5), we get $P(BC)=P(ABC)$.

Ex. 8. *The events $E_1, E_2, \ldots\ldots, E_n$ are mutually exclusive and $E=E_1+E_2+\cdots+E_n$. Show that if $P(A\mid E_i)=P(B\mid E_i)$, $i=1, 2, \ldots\ldots, n$, then $P(A\mid E)=P(B\mid E)$. Is the conclusion true even if the events E_i are not mutually exclusive ?*

[C. H. (Math.) '80]

We have $P(A\mid E)=\dfrac{P(AE)}{P(E)}$. ... (1)

Now $AE=A(E_1+E_2+\cdots+E_n)=AE_1+AE_2+\cdots+AE_n$.

Here $(AE_i)(AE_j)=A(E_iE_j)=AO=O$, E_i, E_j being mutually exclusive for $i, j=1, 2, \ldots, n$; $i\neq j$. This implies that AE_1, $AE_2, \ldots, AE_n$ are pairwise mutually exclusive and hence

$$P(AE)=P(AE_1)+P(AE_2)+\cdots+P(AE_n). \quad \ldots (2)$$

Now $P(AE_i)=P(E_i)P(A\mid E_i)$, $i=1, 2, \ldots, n$.

$$\therefore\ P(AE)=P(E_1)P(A\mid E_1)+P(E_2)P(A\mid E_2)$$
$$+\cdots\cdots+P(E_n)P(A\mid E_n)$$

$$\therefore\ P(E)P(A\mid E)=P(E_1)P(A\mid E_1)+P(E_2)P(A\mid E_2)$$
$$+\cdots\cdots+P(E_n)P(A\mid E_n)$$

$$\therefore \quad P(A \mid E)=\frac{P(E_1)}{P(E)}P(A \mid E_1)+\frac{P(E_2)}{P(E)}P(A \mid E_2)$$
$$+\ldots\ldots+\frac{P(E_n)}{P(E)}P(A \mid E_n),$$

$\therefore$ here $P(E) \neq 0$.

Similarly,
$$P(B \mid E)=\frac{P(E_1)}{P(E)}P(B \mid E_1)+\frac{P(E_2)}{P(E)}P(B \mid E_2)$$
$$+\ldots\ldots+\frac{P(E_n)}{P(E)}P(B \mid E_n),$$

$\therefore$ $P(A \mid E)=P(B \mid E)$, since $P(A \mid E_i)=P(B \mid E_i)$ for $i=1, 2, \ldots\ldots, n$.

The above result will not be true if the events E_i are not mutually exclusive, for, in that case relation (2) will not be true.

Ex. 9. *Prove that* $P(B \mid A) \geq 1-\dfrac{P(\overline{B})}{P(A)}$ *in general.*

[C. H. (Math.) '66]

We have $P(\overline{B})=1-P(B)$.

Now $P(A+B)=P(A)+P(B)-P(AB)$ and $0 \leq P(A+B) \leq 1$.

$\therefore$ $0 \leq P(A)+P(B)-P(AB) \leqslant 1$

or, $1-P(B) \geq P(A)-P(AB)$

or, $P(\overline{B}) \geq P(A)-P(AB)$

Henee $\dfrac{P(\overline{B})}{P(A)} \geqslant 1-\dfrac{P(AB)}{P(A)}=1-P(B \mid A)$.

or, $P(B \mid A) \geq 1-\dfrac{P(\overline{B})}{P(A)}$.

Ex. 10. *If any one of the pairs* (A, B), $(A, \overline{B})$, $(\overline{A}, B)$ *and* $(\overline{A}, \overline{B})$ *is an independent pair of events, then show that all the other pairs are independent pairs.* [C.H. (Math.) '74, '82, '88, '91, '92]

Let A and B be independent.

We have $AB+A\overline{B}=A$ and AB, $A\overline{B}$ are mutually exclusive.

$\therefore$ $P(A)=P(AB)+P(A\overline{B})$

or, $P(A\overline{B})=P(A)-P(A)P(B)=P(A)[1-P(B)]=P(A)P(\overline{B})$,

since A and B are independent.

Thus A and $\overline{B}$ are independent, when A and B are independent. ... (1)

Again $\overline{A}B+AB=B$ and $\overline{A}B$, AB are mutually exclusive.

$\therefore$ $P(\overline{A}B)+P(AB)=P(B)$.

Hence if A and B are independent, then
$$P(\overline{A}B)=P(B)-P(AB)=P(B)-P(A)P(B)$$
$$=P(B)\{1-P(A)\}=P(\overline{A})P(B).$$

Thus $\overline{A}$ and B are independent if A and B are independent. ... (2)

Now let $\overline{A}$ and B be independent, then by (1), $\overline{A}$ and $\overline{B}$ are independent. Again let $\overline{A}$ and $\overline{B}$ be independent. Then by (1) $\overline{A}$ and $\overline{\overline{B}}$, *i.e.*, $\overline{A}$ and B are independent and hence by (2) $\overline{\overline{A}}$ and B, *i.e.*, A and B are independent. Finally from the last result, if $\overline{A}$ and B are independent, then $\overline{\overline{A}}$ and $\overline{B}$, *i.e*, A and $\overline{B}$ are independent. Thus it is proved that if any one of the given pairs is independent, then all the other pairs are independent pairs.

Ex. 11. *Consider events* A *and* B *such that* $P(A)=\frac{1}{4}$, $P(B \mid A)=\frac{1}{2}$, $P(A \mid B)=\frac{1}{4}$. *Find* $P(\overline{A} \mid \overline{B})$ *and* $P(A \mid B)+P(A \mid \overline{B})$.

[C. H. (Math.) '76]

We have $\frac{1}{2}=P(B \mid A)=\dfrac{P(AB)}{P(A)}=4P(AB)$ $\therefore$ $P(AB)=\frac{1}{8}$.

Again $\frac{1}{4}=P(A \mid B)=\dfrac{P(AB)}{P(B)}=\dfrac{1}{8P(B)}$ $\therefore$ $P(B)=\frac{1}{2}$

$\therefore$ $P(AB)=\frac{1}{8}=P(A)P(B)$.

$\therefore$ A and B are independent. Then by Ex. 10, $\overline{A}$, $\overline{B}$ as also A, $\overline{B}$ are independent.

Now $P(\overline{B})=1-P(B)=\frac{1}{2}$. $P(\overline{A})=1-P(A)=\frac{3}{4}$.

$$\therefore \quad P(A \mid \overline{B})=\frac{P(A\overline{B})}{P(\overline{B})}=\frac{P(A)P(\overline{B})}{P(\overline{B})}=P(A)=\tfrac{1}{4},$$

and
$$P(\overline{A} \mid \overline{B})=\frac{P(\overline{A}\overline{B})}{P(\overline{B})}=\frac{P(\overline{A})P(\overline{B})}{P(\overline{B})}=P(\overline{A})=\tfrac{3}{4}.$$

Finally $P(A \mid B)+P(A \mid \overline{B})=\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$.

Ex. 12. *If* $P(ABC)=0$, *then show that*
$$P(X \mid C)=P(A \mid C)+P(B \mid C)$$
where $X=A+B$. [C. H. (Math. '72)]

We have $P(AC+BC)=P(AC)+P(BC)-P(ABC)$.

Assuming $P(C) \neq 0$, we get
$$\frac{P[(A+B)C]}{P(C)}=\frac{P(AC)}{P(C)}+\frac{P(BC)}{P(C)}-\frac{P(ABC)}{P(C)}$$

or, $\dfrac{P(XC)}{P(C)}=\dfrac{P(AC)}{P(C)}+\dfrac{P(BC)}{P(C)}$ $\quad\because\ P(ABC)=0$

or, $P(X \mid C)=P(A \mid C)+P(B \mid C)$.

Ex. 13. *Let the events A and B be such that A can be partitioned into 3 events A_1, A_2, A_3 and B can be partitioned into two events B_1, B_2; if the events A_i and B_j are pairwise independent for all possible values of i and j, then prove that A and B are independent.* [C.H. (Math.) '86]

We have $A=A_1+A_2+A_3$, $A_iA_j=0$, $i\neq j\ (i,j=1,2,3)$,
and $B=B_1+B_2$ where $B_1B_2=O$.

Now we have $P(AB)=P[(A_1+A_2+A_3)(B_1+B_2)]$

$=P(A_1B_1+A_1B_2+A_2B_1+A_2B_2+A_3B_1+A_3B_2)$,

where we note that any two events of A_1B_1, A_1B_2, A_2B_1, A_2B_2, A_3B_1, A_3B_2 are mutually exclusive.

$\therefore\ P(AB)=P(A_1B_1)+P(A_1B_2)+P(A_2B_1)+P(A_2B_2)+P(A_3B_1)+P(A_3B_2)$

$=P(A_1)P(B_1)+P(A_1)P(B_2)+P(A_2)P(B_1)+P(A_2)P(B_2)+P(A_3)P(B_1)+P(A_3)P(B_2)$

since A_i and B_j are pairwise independent for every i, j $(i=1, 2, 3;\ j=1, 2)$

$=[P(A_1)+P(A_2)+P(A_3)].[P(B_1)+P(B_2)]$

$=P(A_1+A_2+A_3)P(B_1+B_2)$

$[\because A_iA_j=O$ for $i\neq j$; $B_1B_2=O]$

$=P(A)P(B)$, which implies that A and B are independent.

Ex. 14. *A missile was fired at a plane on which there are two targets, T_1 and T_2. The probability of hitting T_1 is p_1 and that of hitting T_2 is p_2. It is known that T_2 was not hit. Find the probability that T_1 was hit.* [C.H. (Math.) '92]

We are to find $P(T_1 \mid \overline{T}_2)=\dfrac{P(T_1\overline{T}_2)}{P(\overline{T}_2)}$ where T_i denotes the event 'the target T_i was hit' for $i=1, 2$.

Now $T_1\overline{T}_2+T_1T_2=T_1$ where $T_1\overline{T}_2$ and T_1T_2 are mutually exclusive.

$\therefore\ P(T_1\overline{T}_2)=P(T_1)-P(T_1T_2)=P(T_1)$, since $T_1T_2=O$, an impossible event.

$$\therefore\ P(T_1 \mid \overline{T}_2)=\frac{P(T_1)}{1-P(T_2)}=\frac{p_1}{1-p_2}.$$

Ex. 15. *A secretary writes four letters and the corresponding addresses on envelopes. If he inserts the letters in the envelopes at random irrespective of address, then calculate the probability that all the letters are wrongly placed.*

Let A_i denote the event 'i-th letter is placed in the correct envelope' $(i=1, 2, 3, 4)$. Then the required event is $\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4$. Now

$P(\overline{A}_1\overline{A}_2\overline{A}_3\overline{A}_4)=P(\overline{A_1+A_2+A_3+A_4})$, by De Morgan's law

$=1-P(A_1+A_2+A_3+A_4)$

$=1-[P(A_1)+P(A_2)+P(A_3)+P(A_4)-P(A_1A_2)$
$-P(A_1A_3)-P(A_1A_4)-P(A_2A_3)-P(A_2A_4)$
$-P(A_3A_4)+P(A_1A_2A_3)+P(A_1A_3A_4)$
$+P(A_1A_2A_4)+P(A_2A_3A_4)-P(A_1A_2A_3A_4)]$

$$=1-\left(4\times\frac{\lfloor 3}{\lfloor 4}-6\times\frac{\lfloor 2}{\lfloor 4}+4\times\frac{1}{\lfloor 4}-\frac{1}{\lfloor 4}\right)=\tfrac{3}{8}.$$

(Note that when the first letter is placed in the correct envelope, then the remaining 3 letters can be placed in $\lfloor 3$ ways etc.)

Ex. 16. *A person takes four tests in succession. The probability of his passing the first test is p, that of his passing each succeeding test is p or $\dfrac{p}{2}$ according as he passes or fails the preceding one. He qualifies provided he passes at least three tests. What is the chance of his qualifying?* [C.H. (Math.) '81]

Let A_i denote the event 'passing the i-th test', $i=1, 2, 3, 4$. The person qualifies if any one of the following mutually exclusive events happens:

$A_1A_2A_3\overline{A}_4$, $A_1A_2\overline{A}_3A_4$, $A_1\overline{A}_2A_3A_4$, $\overline{A}_1A_2A_3A_4$, $A_1A_2A_3A_4$.

So the required probability $=P(A_1A_2A_3\overline{A}_4+A_1A_2\overline{A}_3A_4$
$+A_1\overline{A}_2A_3A_4+\overline{A}_1A_2A_3A_4+A_1A_2A_3A_4)$

$$=P(A_1A_2A_3\bar{A}_4)+P(A_1A_2\bar{A}_3A_4)+P(A_1\bar{A}_2A_3A_4)+P(\bar{A}_1A_2A_3A_4)$$
$$+P(A_1A_2A_3A_4)$$

since the events are mutually exclusive.

$$=P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)P(\bar{A}_4 \mid A_1A_2A_3)$$
$$+P(A_1)P(A_2 \mid A_1)P(\bar{A}_3 \mid A_1A_2)P(A_4 \mid A_1A_2\bar{A}_3)$$
$$+P(A_1)P(\bar{A}_2 \mid A_1)P(A_3 \mid A_1\bar{A}_2)P(A_4 \mid A_1\bar{A}_2A_3)$$
$$+P(\bar{A}_1)P(A_2 \mid \bar{A}_1)P(A_3 \mid \bar{A}_1A_2)P(A_4 \mid \bar{A}_1A_2A_3)$$
$$+P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)P(A_4 \mid A_1A_2A_3)$$
$$=p^3(1-p)+p^2(1-p)\frac{p}{2}+p(1-p)\frac{p}{2}p+(1-p)\frac{p}{2}p^2+p^4$$
$$=\tfrac{1}{2}p^3(5-3p).$$

Ex. 17. *A man seeks advice regarding one of two possible courses of action from three advisers, who arrive at their recommendations independently. He follows the recommendations of the majority. The probabilities that the individual advisers are wrong are* 0.1, 0.05, *and* 0.05 *respectively. What is the probability that the man takes incorrect advice?* [C.H. (Math.) '83]

Let A_i denote the event 'i-th adviser gives incorrect advice' $i=1, 2, 3$. Then $P(A_1)=0{\cdot}1$, $P(A_2)=P(A_3)=0{\cdot}05$. Since the man follows the recommendations of the majority, he will take the incorrect advice when any two or all of the three advisers give incorrect advice. Hence the event that 'the man takes incorrect advice' can happen when any one of the following pairwise mutually exclusive events happens :

$$A_1A_2\bar{A}_3,\ A_1\bar{A}_2A_3,\ \bar{A}_1A_2A_3,\ A_1A_2A_3.$$

So the required probability $= P(A_1A_2\bar{A}_3)+P(A_1\bar{A}_2A_3)$
$$+P(\bar{A}_1A_2A_3)+P(A_1A_2A_3)$$
$$=P(A_1)P(A_2)P(\bar{A}_3)+P(A_1)P(\bar{A}_2)P(A_3)+P(\bar{A}_1)P(A_2)P(A_3)$$
$$+P(A_1)P(A_2)P(A_3)$$

since the events are all independent

$$=(0{\cdot}1)(0{\cdot}05)(0{\cdot}95)+(0{\cdot}1)(0{\cdot}95)(0{\cdot}05)+(0{\cdot}9)(0{\cdot}05)^2+(0{\cdot}1)(0{\cdot}05)^2$$
$$=0{\cdot}012.$$

Ex. 18. *An integer is chosen at random from the first* 100 *positive integers. What is the probability that the integer is divisible by* 6 *or* 8 ? [C.H. (Math.) '83]

If n_1 be the total number of numbers lying between 1 and 100 and divisible by 6, then $96=6+(n_1-1)6$, *i.e.*, $n_1=16$; similarly total number of numbers lying between 1 and 100 and divisible by 8 is 12 and that divisible by both 6 and 8, *i.e.*, divisible by 24 is 4. Now if A_1 be the event 'the integer is divisible by 6' and A_2 be the event 'the integer is divisible by 8', then the required probability$=P(A_1+A_2)=P(A_1)+P(A_2)-P(A_1A_2)$

$$=\tfrac{16}{100}+\tfrac{12}{100}-\tfrac{4}{100}=\tfrac{6}{25}.$$

Ex. 19. *With probability* p, *a car travels along the road every second independent of the other time moments. A pedestrian needs three seconds to cross the road. What is the probability that a pedestrian approaching the road will wait* (*i*) 3 *seconds,* (*ii*) 4 *seconds to cross the road ?* [C.H. (Math.) '86]

(*i*) The event that the pedestrian will wait for 3 seconds to cross the road will happen iff while one car crosses the road in each of the first three seconds and in the next 3 seconds no car crosses the road so that the pedestrian gets the requisite time to cross the road.

So the required probability $=p.p.p\times(1-p)(1-p)(1-p)$
$$=p^3(1-p)^3.$$

(*ii*) Similarly probability in this case $=P^4(1-p)^3$.

Ex. 20. *A total number of n shells are fired at a target. The probability of the i-th shell hitting the terget is* p_i, $i=1, 2, \ldots\ldots, n$. *Assuming that the n firings are n mutually independent events, find the probability that at least two shells out of n hit the target.* [C.H. (Math.) '88]

Let A_i denote the event 'exactly i shells hit the target', $i=0, 1, 2, \ldots\ldots, n$. Then, since the n firings are considered as n mutually independent events,

$$P(A_0)=(1-p_1)(1-p_2)\ldots\ldots(1-p_n)$$

and $P(A_1)=\sum_{i=1}^{n}(1-p_1)(1-p_2)\ldots(1-p_{i-1})p_i(1-p_{i+1})\ldots(1-p_n)$

$$=(1-p_1)(1-p_2)\ldots\ldots(1-p_n)\sum_{i=1}^{n}\frac{p_i}{1-p_i}$$

$\therefore$ the required probability $P(\overline{A_0+A_1})$

$$=1-[P(A_0)+P(A_1)]=1-(1-p_1)\cdots(1-p_n)\left[1+\sum_{i=1}^{n}\frac{p_i}{1-p_i}\right]$$

Ex. 21. *On an attempt to land an unmanned rocket on the moon, the probability of a successful landing is 0·4. The probability that monitoring system will give the correct information concerning landing is 0·9 in either case. A shot is made and a successful landing is indicated by the monitoring system. What is the probability of a successful landing?* [C.H. (Math.) '82]

Let the events X, A_1, A_2 be defined as :

X: 'successful landing';

A_1: 'monitoring system indicates successful landing'

A_2: 'monitoring system indicates unsuccessful landing'.

Then $P(X)=0.4$, $P(A_1 \mid X)=0.9$, $P(A_2 \mid \overline{X})=0.9$.

Now $X+\overline{X}=S$, the event space.

$\therefore$ $A_1X+A_1\overline{X}=A_1S=A_1$.

$\therefore$ $P(A_1X)+P(A_1\overline{X})=P(A_1)$, since A_1X and $A_1\overline{X}$ are mutually exclusive events.

or, $P(X)\,P(A_1 \mid X)+P(\overline{X})\,P(A_1 \mid \overline{X})=P(A_1)$

or, $(.4)(.9)+(.6)(.1)=P(A_1)$, since $P(A_1 \mid \overline{X})=1-P(\overline{A}_1 \mid \overline{X})$
$=1-P(A_2 \mid \overline{X})$
$=1-{\cdot}9={\cdot}1$

or, $P(A_1)={\cdot}42$

So the required probability $=P(X \mid A_1)=\dfrac{P(A_1X)}{P(A_1)}$

$$=\frac{P(X)P(A_1 \mid X)}{P(A_1)}=\frac{6}{7}.$$

Ex. 22. *A jar contains two white balls and three black balls. The balls are drawn from the jar one by one and placed on the table in the order drawn. What is the probability that they are drawn in the order white, black, black, white, black?* [C.H. (Math.) '85]

Let A_1, A_2, A_3, A_4, A_5 be the events 'first ball drawn is white', 'second ball drawn is black', 'third ball drawn is black', 'fourth ball drawn is white' and 'fifth ball drawn is black' respectively.

Then $P(A_1)=\frac{2}{5}$, $P(A_2 \mid A_1)=\frac{3}{4}$, $P(A_3 \mid A_1A_2)=\frac{2}{3}$,

$P(A_4 \mid A_1A_2A_3)=\frac{1}{2}$ and $P(A_5 \mid A_1A_2A_3A_4)=1$.

$\therefore$ required probability $=P(A_1A_2A_3A_4A_5)$

$$=P(A_1)P(A_2 \mid A_1)\,P(A_3 \mid A_1A_2)\,P(A_4 \mid A_1A_2A_3)\,P(A_5 \mid A_1A_2A_3A_4)$$

$=0.1$

Ex. 23. *There are two identical urns containing 4 white and 3 red balls; 3 white and 7 red balls. An urn is chosen at random and a ball is drawn from it. Find the probability that the ball is white. What is the probability that it is from the first urn if the ball drawn is white?*

Let A_1, A_2 and X be the events 'first urn chosen', 'second urn chosen' and 'ball drawn is white' respectively.

Then $P(A_1)=P(A_2)=\frac{1}{2}$.

Since $A_1+A_2=S$, the corresponding event space,

$X=SX=(A_1+A_2)\,X=A_1X+A_2X$ and the two events A_1X and A_2X are mutually exclusive.

$$P(X)=P(A_1X)+P(A_2X)=P(A_1)P(X \mid A_1)+P(A_2)P(X \mid A_2)$$

$=\frac{1}{2}\cdot\frac{4}{7}+\frac{1}{2}\cdot\frac{3}{10}=\frac{61}{140}$, which gives result for the first part.

Again, $P(A_1 \mid X)=\dfrac{P(A_1X)}{P(X)}=\dfrac{P(A_1)P(X \mid A_1)}{P(X)}=\dfrac{\frac{1}{2}\times\frac{4}{7}}{\frac{61}{140}}=\dfrac{40}{61}$.

Ex. 24. *From an urn containing 5 white and 5 black balls, 5 balls are transferred at random into an empty second urn from which one ball is drawn and it is found to be white. What is the probability that all balls transferred from the first urn are white?* [C.H. (Math.) '71]

Let E_0, E_1, E_2, E_3, E_4, E_5 be the events defined as follows ;

E_0: '5 black balls drawn from the first urn'

E_1: '1 white and 4 black balls drawn from the first urn'

E_2: '2 white and 3 black balls drawn from the first urn'

E_3: '3 white and 2 black balls drawn from the first urn'

E_4: '4 white and 1 black ball drawn from the first urn'

E_5: '5 white balls drawn from the first urn.'

Let also X be the event one white ball drawn from the second urn.'

Then $P(E_0)=\frac{^5C_5}{^{10}C_5}=\frac{1}{252}$, $P(E_1)=\frac{^5C_1\times{}^5C_4}{^{10}C_5}=\frac{25}{252}$,

$P(E_2)=\frac{^5C_2\times{}^5C_3}{^{10}C_5}=\frac{25}{63}$, $P(E_3)=\frac{^5C_3\times{}^5C_2}{^{10}C_5}=\frac{25}{63}$,

$P(E_4)=\frac{^5C_4\times{}^5C_1}{^{10}C_5}=\frac{25}{252}$, $P(E_5)=\frac{^5C_5}{^{10}C_5}=\frac{1}{252}$.

Also, $P(X \mid E_0)=0$, $P(X \mid E_1)=\frac{1}{5}$, $P(X \mid E_2)=\frac{2}{5}$, $P(X \mid E_3)=\frac{3}{5}$, $P(X \mid E_4)=\frac{4}{5}$ and $P(X \mid E_5)=1$.

Required probability

$$=P(E_5 \mid X)=\frac{P(E_5)P(X \mid E_5)}{\sum_{i=0}^{5}P(E_i)P(X \mid E_i)}, \text{ by Baye's Theorem.}$$

$$=\frac{1}{126}.$$

Ex. 25. *Two persons agree to play a game by drawing balls by turn from a box containing 4 white and 6 black balls. He who draws the first white ball wins. Find the probability that the man who starts the game loses the game.* [C. H. (Math.) '71]

Let A start the game and B be the other player. Let also E_i be the event 'white ball appears when the i-th ball is drawn'.

Now A wins if any one of the following mutually exclusive events happens:

E_1, $\bar{E}_1\bar{E}_2E_3$, $\bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4E_5$, $\bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4\bar{E}_5\bar{E}_6E_7$, since there are 6 black balls in all.

$\therefore$ the probability that A wins

$$=P(E_1)+P(\bar{E}_1\bar{E}_2E_3)+P(\bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4E_5)$$
$$+P(\bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4\bar{E}_5\bar{E}_6E_7). \qquad (1)$$

Now $P(E_1)=\frac{4}{10}$, $P(\bar{E}_1\bar{E}_2E_3)=P(\bar{E}_1)P(\bar{E}_2 \mid \bar{E}_1)P(E_3 \mid \bar{E}_1\bar{E}_2)$

$$=\frac{6}{10}\times\frac{5}{9}\times\frac{4}{8}=\frac{1}{6},$$

$P(\bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4E_5)$

$$=P(\bar{E}_1)P(\bar{E}_2 \mid \bar{E}_1)P(\bar{E}_3 \mid \bar{E}_1\bar{E}_2)P(\bar{E}_4 \mid \bar{E}_1\bar{E}_2\bar{E}_3)$$
$$P(E_5 \mid \bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4)$$

$$=\frac{6}{10}\times\frac{5}{9}\times\frac{4}{8}\times\frac{3}{7}\times\frac{4}{6}=\frac{1}{21},$$

and $P(\bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4\bar{E}_5\bar{E}_6E_7)=P(\bar{E}_1)P(\bar{E}_2 \mid \bar{E}_1)P(\bar{E}_3 \mid \bar{E}_1\bar{E}_2)$

$$P(\bar{E}_4 \mid \bar{E}_1\bar{E}_2\bar{E}_3)P(\bar{E}_5 \mid \bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4)$$
$$P(\bar{E}_6 \mid \bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4\bar{E}_5)P(E_7 \mid \bar{E}_1\bar{E}_2\bar{E}_3\bar{E}_4\bar{E}_5\bar{E}_6)$$

$$=\frac{6}{10}\times\frac{5}{9}\times\frac{4}{8}\times\frac{3}{7}\times\frac{2}{6}\times\frac{1}{5}\times\frac{4}{4}=\frac{1}{210}.$$

Therefore, from (1), the probability that A wins

$$=\frac{4}{10}+\frac{1}{6}+\frac{1}{21}+\frac{1}{210}=\frac{13}{21}.$$

$\therefore$ the probability that the man who starts the game loses $=1-\frac{13}{21}=\frac{8}{21}$.

Ex. 26. *Two players A and B alternately throw a pair of die; A wins if A throws 6 before B throws 7 and B wins if B throws 7 before A throws 6. If A begins, then find the probability that A wins.*

Let A_i denote the event 'A throws 6 in the i th throw' and B_j denote the event 'B throws 7 in the j th throw'. Now A wins if at least one of the following mutually exclusive events happens: A_1, $\bar{A}_1\bar{B}_2A_3$, $\bar{A}_1\bar{B}_2\bar{A}_3\bar{B}_4A_5$, ...

Now the event 'throwing 6 by two dice' contains the 5 distinct outcomes, namely (1, 5), (2, 4), (3, 3), (4, 2), (5, 1) and the event 'throwing 7 by 2 dice' contains the 6 distinct outcomes, namely, (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1).

$\therefore$ $P(A_i)=\frac{5}{36}$, $P(B_j)=\frac{6}{36}=\frac{1}{6}$; $i=1, 3, 5, \ldots\ldots$ and $j=2, 4, 6, \ldots$

Hence the probability that A wins

$$=P(A_1+\bar{A}_1\bar{B}_2A_3+\bar{A}_1\bar{B}_2\bar{A}_3\bar{B}_4A_5+\cdots\cdots)$$
$$=P(A_1)+P(\bar{A}_1\bar{B}_2A_3)+P(\bar{A}_1\bar{B}_2\bar{A}_3\bar{B}_4A_5)+\cdots\cdots$$
$$=\frac{5}{36}+\frac{31}{36}\cdot\frac{5}{6}\cdot\frac{5}{36}+\frac{31}{36}\cdot\frac{5}{6}\cdot\frac{31}{36}\cdot\frac{5}{6}\cdot\frac{5}{36}+\cdots\cdots$$
$$=\frac{5}{36}+\frac{5}{36}\left(\frac{31}{36}\cdot\frac{5}{6}\right)+\frac{5}{36}\left(\frac{31}{36}\cdot\frac{5}{6}\right)^2+\cdots\cdots$$
$$=\frac{\frac{5}{36}}{1-\frac{31}{36}\cdot\frac{5}{6}}=\frac{30}{61}.$$

Ex. 27. *A die is thrown twice, the event space S consists of the 36 possible pairs of outcome (a, b), each assigned with probabilities $\frac{1}{36}$. Let A, B, C denote the events:*

$A\equiv\{(a, b) \mid a \text{ is odd}\}$

$B\equiv\{(a, b) \mid b \text{ is odd}\}$

$C\equiv\{(a, b) \mid a+b \text{ is odd}\}$.

Check whether A, B and C are independent or independent in pairs only. [C. H. (Math.) '81]

Number of simple events favourable to $A=3\times6=18$.

$\therefore$ $P(A)=\frac{18}{36}=\frac{1}{2}$.

Similarly, $P(B)=\frac{1}{2}$, $P(C)=\frac{1}{2}$.

Number of simple events favourable to $AB=3\times 3=9$.

$\therefore\ P(AB)=\frac{9}{36}=\frac{1}{4}=P(A)P(B)$.

Number of simple events favourable to $BC=9$ [In fact the simple events favourable to BC are {(2,1), (4,1), (6,1), (2,3), (4,3), (6,3), (2,5), (4,5), (6,5)}].

$\therefore\ P(BC)=\frac{9}{36}=\frac{1}{4}=P(B)P(C)$.

Similarly $P(CA)=P(C)P(A)$.

Now the event $ABC=\{(a,b) \mid a,b,\ a+b$ are all odd$\}=O$.

$\therefore P(ABC)=0\neq P(A)\,P(B)\,P(C)$.

Thus the events A, B, C are not independent but are pairwise independent only.

Ex. 28. *The probabilities of n independent events are p_1, $p_2,\ldots,p_n$. Find the probability that at least one of the events will happen.* [C.H. (Math.) '63]

Let $A_1, A_2, A_3,\ldots\ldots, A_n$ be the given independent events where $P(A_i)=p_i$ for $i=1,2,\ldots,n$. The required event is $A_1+A_2+\cdots+A_n$. So the required probability is $P(A_1+A_2+\cdots+A_n)$

$$=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\ldots A_n)$$

$$=\sum_{i=1}^{n}p_i-\sum_{i<j}P(A_i)P(A_j)+\sum_{i<j<k}P(A_i)P(A_j)P(A_k)-\cdots$$

$$\cdots+(-1)^{n-1}P(A_1)P(A_2)\ldots P(A_n)$$

since $A_1, A_2,\ldots\ldots, A_n$ are independent

$$=\sum_{i=1}^{n}p_i-\sum_{i<j}p_ip_j+\sum_{i<j<k}p_ip_jp_k-\cdots+(-1)^{n-1}(p_1p_2\ldots\ldots p_n)$$

$$=1-\Big[1-\sum_{i=1}^{n}p_i+\sum_{i<j}p_ip_j-\sum_{i<j<k}p_ip_jp_k+\cdots\cdots+(-1)^np_1p_2\ldots\ldots p_n\Big]$$

$$=1-[(1-p_1)(1-p_2)\ldots\ldots(1-p_n)]$$

$$=1-(1-p_1)(1-p_2)\ldots\ldots(1-p_n).$$

Ex. 29. *Find the probability that in a game of bridge, a hand of 13 cards will contain (i) all the aces, (ii) at last one ace.* [C. H. (Math.) '67]

Total number of simple events in the event space $={}^{52}C_{13}$, as a simple event will be a group of 13 cards.

Now a pack of 52 cards contains 4 aces.

We assume that all simple events are equally likely.

(*i*) The required probability $=\dfrac{{}^4C_4\times{}^{48}C_9}{{}^{52}C_{13}}$.

(*ii*) The event 'no ace' contains ${}^{48}C_{13}$ simple events.

$\therefore$ the probability that a hand will contain no ace

$$=\frac{{}^{48}C_{13}}{{}^{52}C_{13}}.$$

$\therefore$ required probability $=1-\dfrac{{}^{48}C_{13}}{{}^{52}C_{13}}$.

Ex. 30. *7 mathematics and 3 physics books are placed at random on a book-shelf. Find the probability that none of the physics books are placed consecutively.* [C.H. (Math.) '72]

Total number of simple events in the event space $=\underline{|10}$. Let A be the event 'none of the physics books are placed consecutiveiy'.

The number of simple events favourable to A

= number of ways in which 10 books can be arranged where the 3 physics books are placed in between and at the ends of 7 mathematics books (*i.e.*, in 8 places) $={}^8P_3\times\underline{|7}$

Assuming that all simple events are equally likely,

by classical rule required probability $=\dfrac{{}^8P_3\times\underline{|7}}{\underline{|10}}=\dfrac{7}{15}$.

Ex. 31. *One card is selected at random from 100 cards numbered 00, 01,......, 98, 99. Suppose n_1 and n_2 are the sum and product respectively of the digits on the selected card. Find $P\{n_1=i \mid n_2=0\}$.* [C.H. (Math.) '87]

We have $P\{n_1=i \mid n_2=0\}=\dfrac{P\{(n_1=i)\cap(n_2=0)\}}{P(n_2=0)}$.

Case I. Let $i=0$. The outcomes favourable to the event '$n_2=0$' are 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 20, 30, 40, 50, 60, 70, 80, 90, *i.e.*, total number of simple events connected to the event

'$n_2=0$' is 19, of which the only simple event connected to the event '$(n_1=0)\cap(n_2=0)$' is {00}.

Hence $P\{n_1=0 \mid n_2=0\}=\frac{\frac{1}{100}}{\frac{19}{100}}=\frac{1}{19}$.

Case II. Let $i \neq 0$. In this case for any particular $i \neq 0$, only two simple events are connected to the event '$(n=i)\cap(n_2=0)$'. For example, if $i=1$, events {01}, {10} are connected to the event '$(n_1=1)\cap(n_2=0)$', if $i=2$, events {02}, {20} are connected to the event '$(n_1=2)\cap(n_2=0)$' and so on. Hence, $P(n_1=i \mid n_2=0)=\frac{2}{19}$, $i=1, 2, 3, \ldots\ldots, 9$.

Ex. 32. *Consider all families with two children and assume that each child is equally likely to be a girl or a boy. If such a family is picked at random and found to have a boy, then what is the probability that it has another boy?* [C.H. (Math.) '90]

The event space consists of four simple events $\{(b,b), (b, g), (g, b), (g, g)\}$, where '$b$' stands for a boy and '$g$' for a girl. Let A be the event 'the family has a boy', and B be the event that 'the family has another boy'. The second event implies that the family has both the two children boys. We will have to find the probability $P(B \mid A)$.

Now $P(A)=\frac{3}{4}$, $P(AB)=\frac{1}{4}$.

$$\therefore \quad P(B \mid A)=\frac{P)AB)}{P(A)}=\frac{1/4}{3/4}=\frac{1}{3}.$$

Ex. 33. *In an examination 30% of the students failed in mathematics, 15% of the students failed in chemistry and 10% of the students failed in both chemistry and mathematics. A student is selected at random. If he failed in chemistry, then what is probability that he passed in mathematics?* [C.H. (Math.) '89]

Let A and B denote the events 'a student failed in mathematics' and 'a student failed in chemistry' respectively.

Then $P(A)=\frac{3}{10}$, $P(B)=\frac{3}{20}$, $P(AB)=\frac{1}{10}$.

The required probability $=P(\overline{A} \mid B)=1-P(A \mid B)$.

$$=1-\frac{P(AB)}{P(B)}=1-\frac{\frac{1}{10}}{\frac{3}{20}}=\frac{1}{3}.$$

Ex. 34. *The chance that a doctor will diagnose a certain disease correctly is 60%. The chance that a patient will die by his treatment after correct diagnosis is 40% and the chance of death by wrong diagnosis is 70%. A patient of the doctor who had the disease dies. What is the probability that the disease was diagnosed correctly?* [C.H. (Math.) '89]

Let E_1 denote the event 'the disease was diagnosed correctly by the doctor' and E_2 denote the event 'a patient who has the disease dies'. We are given that

$P(E_1)=\frac{3}{5}$, $P(E_2 \mid E_1)=\frac{2}{5}$, $P(E_2 \mid \overline{E}_1)=\frac{7}{10}$.

We are to find $P(E_1 \mid E_2)$. Now the event E_2 can materialize in the following two mutually exclusive ways:

$$E_2E_1 \text{ and } E_2\overline{E}_1$$

$$\therefore \quad P(E_2)=P(E_2E_1)+P(E_2\overline{E}_1)=P(E_1)P(E_2 \mid E_1)+P(\overline{E}_1)P(E_2 \mid \overline{E}_1)$$

$$=\frac{3}{5}\times\frac{2}{5}+\frac{2}{5}\times\frac{7}{10}=\frac{13}{25}.$$

$$\therefore \quad P(E_1 \mid E_2)=\frac{P(E_1E_2)}{P(E_2)}=\frac{P(E_1)P(E_2 \mid E_1)}{P(E_2)}=\frac{\frac{3}{5}\times\frac{2}{5}}{\frac{13}{25}}=\frac{6}{13}.$$

Ex. 35. *Two numbers are selected at random from the set $\{1,2,\ldots, n\}$. What is probability that the difference between the first and second chosen number is $\geq m$, a positive integer?* [C.H. (Math.) '91]

Let x and y be the first and second numbers chosen at random. We are to find $P(\mid x-y \mid \geq m)$. We first note that if $m \geq n$, then the event '$\mid x-y \mid \geq m$' is an impossible event. So we assume that $m < n$.

Now different choices of x and y such that $\mid x-y \mid \geq m$, where $m < n$ can be done as follows:

If one of x and y is 1, the other may be $m+1, m+2, \ldots\ldots, n$; [*i.e.*, $(n-m)$ choices].

If one of x and y is 2, the other may be $m+2, m+3, \ldots\ldots, n$; [*i.e.*, $(n-m-1)$ choices].

... ...

If one of x and y is $n-m$, the other is n [*i.e.*, only one choice].

Thus the total number of choices of x and y such that $|x-y|>m$ is $(n-m)+(n-m-1)+\ldots\ldots+1=\dfrac{(n-m)(n-m+1)}{2}$.

Hence $P(|x-y|>m)=\dfrac{(n-m)(n-m+1)}{2.{}^nC_2}=\dfrac{(n-m)(n-m+1)}{n(n-1)}$, that can be chosen at random as the total number of unordered pairs $\{x, y\}$ that can be chosen at random from the given set $\{1,2,\ldots\ldots,n\}$ is ${}^nC_2$.

Ex. 36. *Find the probability P_N that a natural number chosen at random from the set $\{1,2,\ldots\ldots, N\}$ is divisible by a fixed natural number k. Find limit P_N.* [C.H. (Math.) '87]

Let $[z]$ denote the largest integer contained in z, z being a rational number. Then number of integers in the given set $\{1,2,\ldots\ldots, N\}$ that are divisible by k is $\left[\dfrac{N}{k}\right]$.

[For example $[\frac{17}{5}]=3$ and the number of integers among $\{1,2,\ldots, 17\}$ that are divisible by 5 is 3, namely 5, 10, 15].

Also the event space contains N simple events.

$\therefore$ required probability $P_N=\dfrac{1}{N}\left[\dfrac{N}{k}\right]$.

If $\left[\dfrac{N}{k}\right]=q$, then by division algorithm,

$N=kq+r,\ 0\leqslant r<k.$

$$\therefore\ p_N=\frac{1}{N}\left\{\left[\frac{N}{k}\right]\right\}=\frac{q}{N}=\frac{N-r}{kN}=\frac{1}{k}-\frac{r}{kN}.$$

Now $0\leqslant\dfrac{r}{kN}<\dfrac{1}{N}$. Also $\underset{N\to\infty}{Lt}\ \dfrac{1}{N}=0.$ $\therefore$ $\underset{N\to\infty}{Lt}\ \dfrac{r}{kN}=0.$

$$\therefore\ \underset{N\to\infty}{Lim}\ p_N=\frac{1}{k}.$$

Ex. 37. *Show that the probability of obtaining six at least once in 4 throws with a die is slightly greater than* $0{\cdot}5$.

Total number of simple events in the event space in 6^4. Let A denote the event 'at least one six'. Then the event $\bar{A}$ (*i.e.*, event 'no six') has 5^4 simple events.

$\therefore\ P(\bar{A})=(5/6)^4.$

$\therefore\ P(A)=1-(5/6)^4\simeq{\cdot}52$, which is slightly greater than ${\cdot}5$.

Ex. 38. *Find the minimum number of times a fair die has to be thrown such that the probability of no six is less than $\frac{1}{2}$.*

[C.H. (Math.) '92]

By Ex. 37, probability of the event 'no six' in n throws $=(5/6)^n$.

Now $(\frac{5}{6})^n<\frac{1}{2}$ gives $n\geq 4$.

Hence the minimum number of times the die has to be thrown is 4.

Ex. 39. *The integers x and y are chosen at random with replacement from nine natural numbers $1, 2,\ldots\ldots, 8, 9$. Find the probability that $|x^2-y^2|$ is divisible by 2.* [C.H. (Math.) '93]

x^2-y^2 is divisible by 2, if x, y are both even or both odd. Now total number of pair of numbers, both even, that can be chosen from $\{1, 2,\ldots\ldots, 8, 9\}$ with replacement is 4×4. Similarly total number of pair of numbers, both odd, that can be selected from $\{1, 2,\ldots\ldots, 8, 9\}$ with replacement is 5×5.

Also two integers x and y can be chosen at random with replacement from first nine natural numbers in $9\times9=81$ ways.

Hence the required probability$=\dfrac{16+25}{81}=\dfrac{41}{81}$.

Examples III

1. Let A, B, C be three mutually independent events. Prove that A and $B+C$ are mutually independent. Prove also that $\bar{A}$ and $\bar{B}\bar{C}$ are also independent.

[$P\{A(B+C)\}=P(AB)+P(AC)-P(ABC)$
$\quad=P(A)P(B)+P(A)P(C)-P(A)P(B)P(C)$

and $P(A)P(B+C)=P(A)\{P(B)+P(C)-P(BC)\}$
$\quad=P(A)P(B)+P(A)P(C)-P(A)P(B)P(C)$

since A, B, C are mutually independent, etc. For second part apply Ex. 10 (Illustrative Examples) for $\bar{A}$ and $\overline{B+C}=\bar{B}\bar{C}$ to be independent.]

2. For any two events A and B, prove that
$P(AB) \le P(A) \le P(A+B) \le P(A)+P(B)$ [C.H. (Math.) '77]

[$AB \subseteq A \subseteq A+B$.
$\therefore \quad P(AB) \le P(A) \le P(A+B) \le P(A)+P(B)$.]

3. For any two events A, B, prove that $P(\bar{A}B)=P(B)-P(AB)$
[See Illustrative Examples, Ex. 2].

4. If $AB=O$, then show that $P(A) \le P(\bar{B})$.
[$A=AB+A\bar{B}=A\bar{B}$. $\quad \therefore \quad A \subseteq \bar{B}$, etc.]

5. Given $P(A)=\frac{1}{2}$, $P(B)=\frac{1}{3}$, $P(AB)=\frac{1}{5}$,
find $P(\bar{A})$, $P(\bar{A}B)$, $P(A\bar{B})$, $P(\bar{A}+\bar{B})$ and $P(A+\bar{B})$.
[See Illustrative Examples, Ex. 2 and 3.]

6. For any events A, B, show that
$P(AB)-P(A)P(B)=P(\bar{A})P(B)-P(\bar{A}B)=P(A)P(\bar{B})-P(A\bar{B})$.
[$P(\bar{A}B)=P(B)-P(AB)$, $P(A\bar{B})=P(A)-P(AB)$.
$\therefore \quad P(AB)-P(A)P(B)=P(AB)-\{1-P(\bar{A})\}P(B)$
$=P(\bar{A})P(B)-\{P(B)-P(AB)\}=P(\bar{A})P(B)-P(\bar{A}B)$ etc.]

7. Given $P(A)=\frac{1}{2}$, $P(B)=\frac{7}{8}$, prove that (a) $P(A+B) \ge \frac{7}{8}$, (b) $\frac{3}{8} \le P(AB) \le \frac{1}{2}$.
[$B \subseteq A+B \Rightarrow P(B) \le P(A+B)$. $\quad \therefore \quad P(A+B) \ge P(B)=\frac{7}{8}$.
Again $A \supseteq AB \Rightarrow P(A) \geqslant P(AB) \quad \therefore \quad P(AB) \le P(A)=\frac{1}{2}$.
and $P(A+B)=P(A)+P(B)-P(AB)$.
$\therefore \quad P(AB) = P(A)+P(B)-P(A+B) \ge P(A)+P(B)-1$
$=\frac{1}{2}+\frac{7}{8}-1=\frac{3}{8}$.]

8. Consider families with 3 children. Let A be the event that the family has children of both sexes and B be the event that there is at most one girl. Examine whether the events A and B are independent.
[The event space contains 8 outcomes ;
(bbb), (bgb), (bbg), (bgg), (gbb), (gbg), (ggb), (ggg),
where 'b' stands for a boy and 'g' for a girl.
$P(A)=\frac{6}{8}=\frac{3}{4}$, $P(B)=\frac{4}{8}=\frac{1}{2}$, $P(AB)=\frac{3}{8}=P(A)P(B)$ etc.]

9. An urn contains four tickets marked with numbers 112, 121, 211, 222 and one ticket is drawn at random. Let $A_i (i=1, 2, 3)$ be an event that the i-th digit of a number of a ticket is 1. Discuss the independence of the events A_1, A_2 and A_3.
[$P(A_1)=P(A_2)=P(A_3)=\frac{2}{4}$; $P(A_1A_2)=\frac{1}{4}=P(A_2A_3)=P(A_3A_1)$
$P(A_1A_2A_3)=0$. A_1, A_2, A_3 are pairwise independent but not mutually independent.]

10. An event A is known to be independent of the events B, $B+C$ and BC. Show that A is independent of C.
[$P(AB)+P(AC)-P(ABC)$
$=P\{A(B+C)\}=P(A)P(B+C)=P(A)\{P(B)+P(C)-P(BC)\}$
$=P(A)P(B)+P(A)P(C)-P(A)P(BC)$
$=P(AB)+P(A)P(C)-P(ABC)$
$\therefore \quad P(AC)=P(A)P(C)$ etc].

11. Prove that if A, B and C are random events and if A, B, C are pairwise independent and A is independent of $B+C$, then A, B, C are mutually independent.
[$P\{A(B+C)\}=P(AB+AC)=P(A)P(B)+P(A)P(C)-P(ABC)$
$P\{A(B+C)\}=P(A)P(B+C)$
$=P(A)P(B)+P(A)P(C)-P(A)P(B)P(C)$
$\therefore \; P(ABC)=P(A)P(B)P(C)$].

12. The numbers 1, 2, ..., n are arranged in random orders. What is the probability that the numbers 1 and 2 are always together.

13. Each of the 12 districts has 2 representatives. Find the probability that in a committee 12 representatives chosen at random (a) a given district is represented (b) all districts are represented.

14. From a set of 17 balls marked 1, 2, 3, ..., 16, 17, one ball is drawn at random. What is the chance that its number is either a multiple of 3 or of 7 ? [C. H. (Math.) '62]

[See Illustrative Examples, Ex. 18.]

15. What is the probability that a bridge hand will have a complete suit ?

16. A bag contains 50 tickets numbered 1, 2, 3, ..., 50 of which 5 are drawn at random and arranged in ascending order of their numbers : $x_1 < x_2 < x_3 < x_4 < x_5$. What is the probability that $x_3 = 30$?

[$x_3=30$, $1 \leq x_1, x_2 \leq 29$ and $31 \leq x_4, x_5 \leq 50$.

17. Two urns contain respectively 4 white, 8 red, 20 black balls and 12 white, 8 red and 12 black balls. One ball is drawn from each urn. Find the probability that both the balls are of the same colour.

18. Four boxes and four balls are numbered 1, 2, 3, 4. The balls are put inside the boxes at random, one ball in each box. What is the probability that at least one ball is put inside a box bearing the same number as the ball ? [C. H. (Math.) '68]

[Required probability $= \dfrac{{}^4C_1 \times 2}{4!} + \dfrac{{}^4C_2}{4!} + \dfrac{1}{4!} = \dfrac{5}{8}$.

19. A box contains 6 red, 4 white and 5 black balls. A person draws 4 balls from the box at random. Find the probability that among the balls there is at least one ball of each colour.

$$\left[\frac{{}^6C_1 \times {}^4C_1 \times {}^5C_2}{{}^{15}C_4} + \frac{{}^6C_1 \times {}^4C_2 \times {}^5C_1}{{}^{15}C_4} + \frac{{}^6C_2 \times {}^4C_1 \times {}^5C_1}{{}^{15}C_4} = 0{\cdot}528. \right]$$

20. Three urns contain respectively a_1 white and b_1 black balls, a_2 white and b_2 black balls and a_3 white and b_3 black balls. One urn is chosen at random and a ball is drawn from it at random. What is the probability that the ball drawn is white ?

[See Illustrative Examples. Ex. 23.]

21. The numbers X, Y are chosen at random from a set of natural numbers $\{1, 2, ..., N\}$, $N \geqslant 3$, with replacement. Find the probability that $| X^2 - Y^2 |$ is divisible by 3.

[See Illustrative Examples Ex. 39.]

22. What is the probability that in a throw of 12 dice each face occurs twice ?

Each die may show any one of the six faces. Hence total number of simple events in the event space $= 6^{12}$.

Also number of simple events connected to the given event = number of ways in which 12 dice may be arranged in 6 groups each of size 2, first group of two dice each showing 1, second group of two dice each showing 2, third group of two dice each showing 3 and so on.]

23. Find the probability that in 5 tossings a coin turns up head at least 3 times in succession.

[We get at least $3H$ in succession if either (*i*) there are no tails, or (*ii*) only $1T$ in the positions first or second or fourth or fifth, or (*iii*) 2 tails in the positions first and second or first and fifth or fourth and fifth.]

24. Two cards are drawn from a well-shuffled pack. Find the probability that at least one of them is spade.

[$A \equiv$ 'first card spade', $B \equiv$ 'second card spade'. Required probability $= P(A\bar{B} + \bar{A}B + AB) = P(A\bar{B}) + P(\bar{A}B) + P(AB)$

$= \frac{13}{52} \cdot \frac{39}{51} + \frac{39}{52} \times \frac{13}{51} + \frac{13}{52} \times \frac{12}{51} = \frac{15}{34}$.]

25. In a survey report it is stated that 66% of male college students in Calcutta like soccer, 37% like hockey and 42% cricket, 11% both soccer and hockey, 10% both soccer and cricket and 12% both hockey and cricket, 8% all three. Show that the statement, as it stands, must be incorrect.

[$P(H+S+C) = P(H) + P(S) + P(C) - P(HS) - P(SC) - P(CH) + P(HSC)$

$= \frac{37}{100} + \frac{66}{100} + \frac{42}{100} - \frac{11}{100} - \frac{10}{100} - \frac{12}{100} + \frac{8}{100} = 1{\cdot}2 > 1.$]

26. A survey regarding liking for novel, comics and poetry among students of a college is made. The survey shows that 20% read novels, 16% read comics and 14% read poetry, 8% read both novel and comics, 6% read novel and poetry and 4% read comics and poetry and 2% read all three. If a student is chosen at random, then find the probability that (*i*) he reads none (*ii*) he reads only one of them.

[(*i*) Required probability $= 1 - P(A+B+C) = 0{\cdot}66$.

(*ii*) Required probability $= P(A\bar{B}\bar{C} + \bar{A}\bar{B}C + \bar{A}B\bar{C})$

$= P(A)P(\bar{B})P(\bar{C}) + P(\bar{A})P(\bar{B})P(C) + P(\bar{A})P(B)P(\bar{C})$

$= {\cdot}35$,

where $A \equiv$ 'a student reads novel', $B \equiv$ 'a student reads comics', $C \equiv$ 'a student reads poetry'.]

27. One shot is fired from each of 3 guns; E_1, E_2, E_3 denote respectively the events that the target is hit by first, second and third gun respectively. If $P(E_1)=\cdot 5$, $P(E_2)=\cdot 6$ and $P(E_3)=\cdot 8$, and if E_1, E_2, E_3 are independent events, then find the probability that (*a*) exactly one hit is registered, (*b*) at least two hits are registered.

[(*a*) Required probability $=P(E_1\bar{E}_2\bar{E}_3)+P(\bar{E}_1E_2\bar{E}_3)$
$+P(\bar{E}_1\bar{E}_2E_3)=\cdot 26.$

(*b*) Required probability $=P(E_1E_2\bar{E}_3)+P(E_1\bar{E}_2E_3)$
$+P(\bar{E}_1E_2E_3)+P(E_1E_2E_3)=\cdot 70.$]

28. A problem in statistics is given to three students A, B and C whose chances of solving it are $\frac{1}{2}$, $\frac{3}{4}$ and $\frac{1}{4}$ respectively. What is the probability that the problem is solved ?

[$E_1\equiv$ 'the problem is solved by A', $E_2\equiv$ 'the problem is solved by B' and $E_3\equiv$ 'the problem is solved by C.

Required probability $=P(E_1+E_2+E_3)=P(E_1)+P(E_2)+P(E_3)$
$-P(E_1E_2)-P(E_2E_3)-P(E_3E_1)+P(E_1E_2E_3)=\frac{29}{32}$.]

29. The odds against A solving a certain problem are 4 to 3 and odds in favour of B solving the problem is 7 to 5. What is the probability that the problem is solved if they both try independently ?

[$A_1\equiv$ 'A solves the problem', $A_2\equiv$ 'B solves the problem'.

$P(A_1)=\frac{3}{7}$, $P(A_2)=\frac{7}{12}$. Required probabiliiy$=P(A_1+A_2)$
$=1-P(\overline{A_1+A_2})=1-P(\bar{A}_1\bar{A}_2)=1-P(\bar{A}_1)P(\bar{A}_2)=\frac{16}{21}$.]

30. (*a*) If A and B are mutually exclusive and $P(B)>0$, then prove that $P(A \mid B)=0$.

(*b*) If A, B and C are any three events, then prove that
$P(A+B \mid C)=P(A \mid C)+P(B \mid C)-P(AB \mid C)$
where $P(C)\neq 0$.

[(*a*) $AB=O$, $\therefore$ $P(AB)=0 \Rightarrow P(A \mid B)=\frac{P(AB)}{P(B)}=0$,

(*b*) $P(A+B \mid C)=\frac{P\{(A+B)C\}}{P(C)}=\frac{P(AC)+P(BC)-P(ABC)}{P(C)}$
$=P(A \mid C)+P(B \mid C)-P(AB \mid C)$.]

31. For any three events A, B, C prove that
$$P(A\bar{B} \mid C)+P(AB \mid C)=P(A \mid C).$$

[Since $A\bar{B}C$ and ABC are mutually exclusive,
$$P(A\bar{B} \mid C)+P(AB \mid C)=\frac{P(A\bar{B}C)+P(ABC)}{P(C)}=\frac{P(A\bar{B}C+ABC)}{P(C)}$$
$$=\frac{P(AC)}{P(C)}=P(A \mid C) \Big].$$

32. For any three events A, B, C such that $B\subset C$, $P(A)>0$, prove that $P(B \mid A)<P(C \mid A)$.

$$\Big[P(C \mid A)=\frac{P(CA)}{P(A)}=\frac{P(BCA+\bar{B}CA)}{P(A)}=P(BC \mid A)+P(\bar{B}C \mid A)$$
$=P(B \mid A)+P(\bar{B}C \mid A)$, since $B\subset C$.
$\therefore$ $P(C \mid A) > P(B \mid A)$.]

33. If A and B are two events and $P(B)\neq 1$, then prove that $P(A \mid \bar{B})=\frac{P(A)-P(AB)}{1-P(B)}$. Hence show that $P(AB)>P(A)+P(B)-1$.

Also show that $P(A)>$ or $<P(A \mid B)$ according as $P(A \mid \bar{B})>$ or $<P(A)$.

[$A\bar{B}+AB=A$. $\therefore$ $P(A \mid \bar{B})=\frac{P(A)-P(AB)}{1-P(B)}$, $\because$ $A\bar{B}$ and AB are mutually exclusive.

Also $P(A\bar{B})=P(A)P(\bar{B} \mid A)=P(\bar{B})P(A \mid \bar{B})$.

$$\therefore \quad \frac{P(A \mid \bar{B})}{P(A)}=\frac{P(\bar{B} \mid A)}{P(\bar{B})}=\frac{1-P(B \mid A)}{1-P(B)}.$$

$\therefore$ $P(A \mid \bar{B}) \lesssim P(A)$ according as $1-P(B \mid A) \lesssim 1-P(B)$

i.e., according as $\frac{P(B \mid A)}{P(B)} \gtrsim 1$

i.e., according as $\frac{P(A \mid B)}{P(A)} \gtrsim 1$,

$\because$ $P(AB)=P(A)P(B \mid A)=P(B)P(A \mid B)$

i.e., according as $P(A)<$ or $>P(A \mid B)$].

34. An urn contains 7 red and 6 black balls. Two balls are drawn without replacement. What is the probability that the second ball is red if it is known that the first is red ?

[$A \equiv$ 'first ball red', $B \equiv$ 'second ball red'.

$P(A)=\frac{7}{13}$, $P(AB)=\frac{{}^{7}P_2}{{}^{13}P_2}$. $\therefore$ $P(B \mid A)=\frac{P(AB)}{P(A)}=\frac{1}{2}$.]

35. Two unbiased dice are thrown. Find the conditional probability that two fives occur if it is known that total is divisible by 5.

[$A \equiv$ 'total is divisible by 5'. $B \equiv$ 'two fives occur.'

A contains 7 simple events—(1, 4), (4, 1), (3, 2), (2, 3), (4, 6), (6, 4), (5, 5).

$$P(B/A)=\frac{P(AB)}{P(A)}=\frac{\frac{1}{36}}{\frac{7}{36}}=\frac{1}{7}.\ ]$$

36. In an office 40% employees have scooters and 20% have cars. Among those who have scooters 80% do not have cars. What is the probability that an employee has a car given that he does not have a scooter?

[$S \equiv$ 'an employee has a scooter', $C \equiv$ 'an employee has a car.'

$P(S)=\cdot 4$, $P(C)=\cdot 2$, $P(\bar{C} \mid S)=\cdot 8$.

$\therefore$ $P(\bar{C}S)=P(S)P(\bar{C} \mid S)=\cdot 32$.

Now $P(\bar{C}S)+P(\bar{C}\bar{S})=P(\bar{C})=\cdot 8$. $\therefore$ $P(\bar{C}\bar{S})=\cdot 8-\cdot 32=\cdot 48$.

Finaly, $P(C\bar{S})+P(\bar{C}\bar{S})=P(\bar{S})=\cdot 6$. $\therefore$ $P(C\bar{S})=\cdot 6-\cdot 48=\cdot 12$.

$\therefore$ $P(C \mid \bar{S})=\frac{P(C\bar{S})}{P(\bar{S})}=\cdot 2$.]

37. A and B are two independent witnesses in a case. The probability that A will speak the truth is x and the probability that B will speak the truth is y. A and B agree in a certain statement. Show that the probability that this statement is true is

$$\frac{xy}{1-x-y+2xy}.$$

[$A_1 \equiv$ 'A and B both speak the truth', $A_2 \equiv$ 'A and B both speak the untruth', $X \equiv$ 'A and B agree in a statement.'

We observe that $X=A_1+A_2$ and A_1, A_2 are mutually exclusive. Then $P(X)=P(A_1)+P(A_2)$.

So $P(X)=x \cdot y+(1-x)(1-y)$.

$$\therefore\ P(A_1 \mid X)=\frac{P(A_1X)}{P(X)}=\frac{P\{A_1(A_1+A_2)\}}{P(X)}=\frac{P(A_1)}{P(X)}=\frac{x \cdot y}{1-x-y+2xy}.\ ]$$

38. An urn contains 15 white and 5 black balls, another urn contains 7 white and 8 black balls. Two balls are transferred from the first urn and placed in the second urn. Then one ball is taken from the latter. What is the probability that it is a white ball?

[Let A_1, A_2, A_3 and X be the events '2 white balls transferred', '2 black balls transferred', '1 white and 1 black balls transferred' and 'one white ball drawn from the second urn' respectively. Required probability

$$=P(A_1X+A_2X+A_3X)$$
$$=P(A_1)P(X \mid A_1)+P(A_2)P(X \mid A_2)+P(A_3)P(X \mid A_3)$$
$$=\frac{{}^{15}C_2}{{}^{20}C_2}\times\frac{9}{17}+\frac{{}^{5}C_2}{{}^{20}C_2}\times\frac{7}{17}+\frac{{}^{15}C_1\times{}^{5}C_1}{{}^{20}C_2}\times\frac{8}{17}=\cdot 5.\ ]$$

39. Two urns contain respectively 5 white and 3 black balls and 4 white and 2 black balls respectively. One ball is transferred from the first to the second urn and then a ball is drawn from the second urn. What is the probability that the ball drawn is white?

[C. H. (Math.) '63]

40. Three urns contain respectively 1 white and 2 black balls, 2 white and 1 black balls, 2 white and 2 black balls. One ball is transferred from the first to the second urn, then one ball is transferred from the second to the third urn, finally one ball is drawn from the third urn. Find the probability that the ball is white.

[$\left.\begin{array}{l} W_i \equiv \text{'a white ball drawn from the } i\text{-th urn'} \\ B_i \equiv \text{'a black ball drawn from the } i\text{-th urn'} \end{array}\right\}$ $i=1, 2, 3.$

$$P(W_2)=P(W_1)P(W_2 \mid W_1)+P(B_1)P(W_2 \mid B_1)$$
$$=\tfrac{1}{3}\cdot\tfrac{3}{4}+\tfrac{2}{3}\cdot\tfrac{2}{4}=\tfrac{7}{12}.$$
$$P(B_2)=1-\tfrac{7}{12}=\tfrac{5}{12}$$
$$P(W_3)=P(W_2)P(W_3 \mid W_2)+P(B_2)P(W_3 \mid B_2)$$
$$=\tfrac{7}{12}\cdot\tfrac{3}{5}+\tfrac{5}{12}\cdot\tfrac{2}{5}=\tfrac{31}{60}.\ ]$$

41. What is the probability that each of the four players, in a bridge game, will get a complete suit of cards?

[$A_i \equiv$ 'ith player will get a complete suit of cards'.

Required probability

$= P(A_1A_2A_3A_4) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)P(A_4 \mid A_1A_2A_3)$

$= \frac{4}{^{52}C_{13}} \times \frac{3}{^{39}C_{13}} \times \frac{2}{^{26}C_{13}} \times \frac{1}{^{13}C_{13}} = \frac{24 \cdot (13!)^4}{52!}$. See Ex. 15].

42. Three urns of the same appearance have the following proportion of white and black balls:

Urn I : 1 white and 2 black balls.

Urn II : 2 white and 1 black balls.

Urn III : 2 white and 2 black balls.

One of the urns is selected and one ball is drawn. It is found to be white. What is the probability that it comes from urn III ?

[See Illustrative Examples, Ex. 24.]

43. In a bolt factory, machines A, B, C manufacture 25, 35 and 40 p.c. of the total respectively. Of this output 5, 4 and 2 p.c. are defective bolts. A bolt is drawn at random from the product and is found defective. What is the probability that it was manufactured by A, B and C respectively ?

[Let A_1, A_2 and A_3 be the events that a bolt is manufactured by A, B and C respectively and E be the event that "a bolt is defective". Then $P(A_1) = \frac{1}{4}$, $P(A_2) = \frac{7}{20}$, $P(A_3) = \frac{2}{5}$.

$P(E \mid A_1) = \frac{1}{20}$, $P(E \mid A_2) = \frac{1}{25}$, $P(E \mid A_3) = \frac{1}{50}$.

Then $P(A_1 \mid E) = \dfrac{P(A_1)P(E \mid A_1)}{\sum_{i=1}^{3} P(A_i)P(E \mid A_i)} = \dfrac{25}{69}$; $P(A_2 \mid E) = \dfrac{28}{69}$.

$P(A_3 \mid E) = \dfrac{16}{69}$.]

44. There are 3 coins, identical in appearance, one of which is ideal and the other two biased with probability $\frac{1}{3}$ and $\frac{2}{3}$ respectively for a head. One coin is taken at random and tossed twice. If a head appears both times, then what is the probability that the ideal coin was chosen ?

[$A_1 \equiv$ 'the ideal coin was chosen'.

$A_2 \equiv$ 'the biased coin, with probability $\frac{1}{3}$ for a head was chosen.'

$A_3 \equiv$ 'the biased coin, with probability $\frac{2}{3}$ for a head was chosen'.

$A \equiv$ 'two heads by two tosses by the chosen coin'.

$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$, $P(A \mid A_1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$,

$P(A \mid A_2) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$.

$P(A \mid A_3) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$.

Required probability $= P(A_1 \mid A) = \dfrac{P(A_1)P(A \mid A_1)}{\sum_{i=1}^{3} P(A_i)P(A \mid A_i)} = \dfrac{9}{29}$.]

45. Urn I contains 2 white, 1 black and 3 red balls; Urn II contains 3 white, 2 black and 4 red balls; Urn III contains 4 white, 3 black and 2 red balls. One urn is chosen at random and 2 balls are drawn. They happen to be red and black. What is the probability that both balls come from urn I ?

[$B_i \equiv$ 'i-th urn chosen', $i = 1, 2, 3$.

$A \equiv$ 'two balls drawn found to be red and black'.

$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$.

$P(A \mid B_1) = \frac{3}{6} \cdot \frac{1}{5} + \frac{1}{6} \cdot \frac{3}{5}$ (first red and second black, first black and second red)

$= \frac{1}{5}$.

$P(A \mid B_2) = \frac{4}{9} \cdot \frac{2}{8} + \frac{2}{9} \cdot \frac{4}{8} = \frac{2}{9}$, $P(A \mid B_3) = \frac{2}{9} \cdot \frac{3}{8} + \frac{3}{9} \cdot \frac{2}{8} = \frac{1}{6}$.

Required probability $= P(B_1 \mid A) = \dfrac{P(B_1)P(A \mid B_1)}{\sum_{i=1}^{3} P(B_i)P(A \mid B_i)} = \dfrac{18}{53}$.]

46. Two urns contain respectively 3 white and 2 black balls and 2 white and 6 black balls. One ball is transferred from urn 1 to urn 2 and then 1 ball is drawn from the latter. It happens to be white. What is the probability that the transferred ball was black ?

[$A_1 \equiv$ 'transferred ball is white', $A_2 \equiv$ 'transferred ball is black' and $B \equiv$ 'ball drawn from the second urn is white'.

Then $P(A_2 \mid B) = \dfrac{P(A_2)P(B \mid A_2)}{\Sigma P(A_i)P(B \mid A_i)} = \frac{4}{13}$, since $P(A_1) = \frac{3}{5}$,

$P(A_2) = \frac{2}{5}$, $P(B \mid A_1) = \frac{3}{9}$, $P(B \mid A_2) = \frac{2}{9}$.]

47. From a vessel containing 3 white and 5 black balls, 4 balls are transferred into an empty vessel. From this vessel a ball is drawn and it is found to be white. What is the probability that out of four balls transferred 3 are white and 1 black.

[$E_i \equiv$ 'i white balls are transferred to the empty vessel', $i=0,1,2,3$.

$B \equiv$ 'one white ball is drawn from the second urn'.

$P(E_0)=\frac{{}^5C_4}{{}^8C_4}$, $P(E_1)=\frac{{}^3C_1 \times {}^5C_3}{{}^8C_4}$, $P(E_2)=\frac{{}^3C_2 \times {}^5C_2}{{}^8C_4}$,

$P(E_3)=\frac{{}^3C_3 \times {}^5C_1}{{}^8C_4}$.

$P(B \mid E_0)=0$, $P(B \mid E_1)=\frac{1}{4}$, $P(B \mid E_2)=\frac{3}{4}$, $P(B \mid E_3)=\frac{3}{4}$.

$$\therefore \quad P(E_3 \mid B)=\frac{P(E_3)P(B \mid E_3)}{\sum_{i=1}^{3} P(E_i)P(B \mid E_i)}=\frac{1}{7}.\ ]$$

48. The population of Cyprus is 75% Greek, 25% Turkish. 20% of the Greeks and 10% of the Turks speak English. A visitor to the town meets someone who speaks English. What is the probability that he is a Greek ?

[$G \equiv$ 'a person is Greek', $T=$'a person is Turkish'

$A \equiv$ 'a person speaks English'.

$P(G)=\frac{3}{4}$, $P(T)=\frac{1}{4}$; $P(A \mid G)=\frac{1}{5}$, $P(A \mid T)=\frac{1}{10}$.

$$\therefore \quad P(G \mid A)=\frac{P(G)P(A \mid G)}{P(G)P(A \mid G)+P(T)P(A \mid T)}=0{\cdot}85.\ ]$$

49. Suppose that there is a chance for a newly constructed house to collapse whether the design is faulty or not. The chance that the design is faulty is 10%. The chance that the house collapses if the design is faulfy is 95% and otherwise it is 45%. It is seen that a house collapsed. What is the probability that it is due to faulty design ?

[$A \equiv$ 'Design is faulty', $B \equiv$ 'the house collapses'.

$P(A)={\cdot}1$, $P(\bar{A})={\cdot}9$, $P(B \mid A)={\cdot}95$, $P(B \mid \bar{A})={\cdot}45$.

$\therefore$ required probability$=P(A \mid B)$

$$=\frac{P(A)P(B \mid A)}{P(A)P(B \mid A)+P(\bar{A})P(B \mid \bar{A})}={\cdot}19.\ ]$$

50. From an urn containing 3 white and 5 black balls, 4 balls are transferred into an empty urn. From the second urn 2 balls are drawn and they happen to be white. What is the probability that the third ball drawn from the same urn will be white ?

$A_1 \equiv$ '2 white and 2 black balls are transferred to the second urn'.

$A_2 \equiv$ '3 white and 1 black balls are transferred to the second urn'.

$A_3 \equiv$ '1 white and 3 black balls are transferred to the second urn.'

$A_4 \equiv$ '4 black balls are transferred to the second urn.'

$A \equiv$ 'both balls drawn from the second urn are white'.

$B \equiv$ 'third ball drawn from the second urn is white'.

$$P(A_1)=\frac{{}^3C_2 \times {}^5C_2}{{}^8C_4}=\frac{3}{7}, \quad P(A_2)=\frac{{}^3C_3 \times {}^5C_1}{{}^8C_4}=\frac{1}{14}.$$

Now we can write $A=A(A_1+A_2+A_3+A_4)$, where A_1, A_2, A_3, A_4 are pairwise mutually exclusive events and $A_1+A_2+A_3+A_4=S$. Now we note that AA_3, AA_4 are both impossible events.

So, $A=A(A_1+A_2)=AA_1+AA_2$ where, $(AA_1)(AA_2)=O$.

$$P(A)=P(AA_1+AA_2)=P(A_1)P(A \mid A_1)+P(A_2)P(A \mid A_2)$$

$$=\frac{3}{7}\cdot\frac{1}{{}^4C_2}+\frac{1}{14}\cdot\frac{{}^3C_2}{{}^4C_2}=\frac{3}{28}.$$

By (3.10.4), $$P(B \mid A)=\frac{\sum_{i=1}^{2} P(A_i)P(A \mid A_i)P(B \mid AA_i)}{\sum_{i=1}^{2} P(A_i)P(A \mid A_i)}.$$

Case I. Let the 2 balls be replaced before the second draw. Then $P(A \mid A_1)=\frac{1}{6}$, $P(A \mid A_2)=\frac{1}{2}$, $P(B \mid AA_1)=\frac{1}{2}$, $P(B \mid AA_2)=\frac{3}{4}$.

$$\therefore \quad P(B \mid A)=\frac{7}{12}.$$

Case II. Let the two balls be not replaced before the second draw. Then $P(B \mid AA_1)=0$, $P(B \mid AA_2)=\frac{1}{2}$. $\therefore$ $P(B \mid A)=\frac{1}{6}$.]

51. Urn A contains 5 black balls and 6 white balls and urn B contains 8 black balls and 4 white balls. Two balls are transferred from B to A and then a ball is drawn from A.

(a) What is the probability that the ball is white?

(b) Given that the ball is white, what is the probability that at least one white ball was transferred to A?

[$B_1 \equiv$ '2 white ball are transferred from B to A'.

$B_2 \equiv$ '1 white ball and 1 black ball are transferred from B to A'.

$B_3 \equiv$ '2 black balls are transferred from B to A'.

$$P(B_1)=\frac{{}^4C_2}{{}^{12}C_2},\ P(B_2)=\frac{{}^8C_1\times{}^4C_1}{{}^{12}C_2},\ P(B_3)=\frac{{}^8C_2}{{}^{12}C_2}.$$

(a) $A\equiv$'one white ball drawn from A'.

$$P(A\mid B_1)=\tfrac{8}{13},\ P(A\mid B_2)=\tfrac{7}{13},\ P(A\mid B_3)=\tfrac{6}{13}.$$

$$P(A)=\sum_{i=1}^{3}P(B_i)P(A\mid B_i)=\frac{20}{39}$$

(b) Required probability $=P(B_1\mid A)+P(B_2\mid A)$

$$=\frac{P(B_1)P(A\mid B_1)}{\sum_{i=1}^{3}P(B_i)P(A\mid B_i)}+\frac{P(B_2)P(A\mid B_2)}{\Sigma P(B_i)P(A\mid B_i)}=\frac{34}{35}.\ ]$$

52. From the numbers $1,2,\ldots\ldots,2n+1$, three are chosen at random. What is the probability that these numbers are in $A.P.$?

[d=common difference of the numbers chosen.

When $d=1$, possible selections are

$(1,2,3),\ (2,3,4),\ldots\ldots,\ (2n-1,\ 2n,\ 2n+1)$

i.e., $(2n-1)$ groups in all.

When $d=2$, possible selections are

$(1,3,5),\ (2,4,6),\ldots\ldots,\ (2n-3,\ 2n-1,\ 2n+1)$

i.e., $(2n-3)$ groups in all.

... ...

When $d=n-1$, possible selections are

$(1,\ n,\ 2n-1),\ (2,\ n+1,\ 2n),\ (3,\ n+2,\ 2n+1)$

i.e., 3 groups in all.

When $d=n$, there is only one possible selection, namely,

$(1,\ n+1,\ 2n+1)$.

Thus total number of selection of groups of 3 numbers connected to the given event $=(2n-1)+(2n-3)+\cdots+5+3+1$

$$=\frac{n}{2}\ (1+2n-1)=n^2.$$

$\therefore$ required probability $=\dfrac{n^2}{{}^{2n+1}C_3}=\dfrac{3n}{4n^2-1}.$]

53. From an urn containing n balls any number of balls are drawn. Show that the probability of drawing an even number of balls is $\dfrac{2^{n-1}-1}{2^n-1}$.

[Total number of event points in the event space

$$={}^nC_1+{}^nC_2+\cdots\cdots+{}^nC_n=2^n-1.$$

The event 'drawing an even number of balls' contains

${}^nC_2+{}^nC_4+\cdots\cdots+{}^nC_r$ simple events,

where $r=\begin{cases} n, \text{ if } n \text{ is even} \\ n-1, \text{ if } n \text{ is odd'}\end{cases}$

i.e., $2^{n-1}-1$ simple events, etc.]

54. From a pack of 52 cards an even number of cards are drawn. Show that the probability that these consists of half red and half black is

$$\frac{\dfrac{52\,!}{(26\,!)^2}-1}{2^{51}-1}$$ [*C.H.* (*Math.*) '62]

[Total number of simple events in the event space

$$={}^{52}C_2+{}^{52}C_4+\cdots\cdots+{}^{52}C_{52}=2^{52-1}-1=2^{51}-1.$$

Now r red and r black cards can be drawn in ${}^{26}C_r\times{}^{26}C_r$ ways (there are 26 red and 26 black cards in all).

$\therefore$ total number of ways in which an even number of cards with half black and half red can be drawn

$$={}^{26}C_1\times{}^{26}C_1+{}^{26}C_2\times{}^{26}C_2+\cdots\cdots+{}^{26}C_{26}\times{}^{26}C_{26}$$

$$=\frac{52\,!}{(26\,!)^2}-1.$$

$\therefore$ required probability $=\left\{\dfrac{52\,!}{(26\,!)^2}-1\right\}\Big/(2^{51}-1.)$

Remember that if $C_0, C_1, \ldots\ldots, C_n$ are the binomial coefficients in the expension of $(1+x)^n$, then

$$C_2+C_4+\cdots\cdots+C_r=2^{n-1}-1 \text{ where } r=\begin{cases} n-1, \text{ if } n \text{ is odd} \\ n, \text{ if } n \text{ is even.}\end{cases}$$

$$C_1^2+C_2^2+\cdots\cdots+C_n^2=\frac{(2n)\,!}{(n\,!)^2}-1.]$$

55. Four dice are thrown. Find the probability that the sum of the numbers will be 18. [C.H. (Math.) '66]

[Total number of simple events in the event space $=6^4$. Now obtaining a sum of 18 from the face values of 4 dice is the same as obtaining a solution of the equation

$$a_1+a_2+a_3+a_4=18 \qquad (1)$$

where a_1, a_2, a_3, a_4 are variables taking values from the set

$$\{1, 2, 3, 4, 5, 6\}.$$

Now if we multiply $(x+x^2+x^3+x^4+x^5+x^6)$ four times by itself, the product will contain terms like

$$x^{a_1+a_2+a_3+a_4}.$$

Hence the total number of distinct solutions of (1) is equal to the coefficient of x^{18} in the expansion of $(x+x^2+x^3+x^4+x^5+x^6)^4$.

Now $(x+x^2+x^3+x^4+x^5+x^6)^4$

$$=\frac{x^4(1-x^6)^4}{(1-x)^4}=x^4(1-x^6)^4(1-x)^{-4}$$

$$=x^4(1-4x^6+6x^{12}-\ldots)\times$$

$$\left\{1+4x+10x^2+\cdots+\frac{n(n+1)(n+2)}{6}x^{n-1}+\cdots\right\}.$$

$\therefore$ coefficient of $x^{18}=\dfrac{15.16.17}{6}-4.\dfrac{9.10.11}{6}+6.\dfrac{3.4.5}{6}$

$=80.$

$\therefore$ Required probability $=\dfrac{80}{6^4}=\dfrac{5}{81}$.]

56. A room has 3 lamp-sockets. From a collection of 10 bulbs of which 6 are bad, a person selects 3 at random and puts them in the sockets. What is the probability that there will be light? [C. H. (Econ.) '90]

[A = 'the room is lighted', then $\bar{A}$ = 'room is dark'. $\bar{A}$ happens, when none of 3 bulbs chosen at random are good.

$$P(\bar{A})=\frac{{}^6C_3}{{}^{10}C_3}=\tfrac{1}{6}. \qquad \therefore\ P(A)=\tfrac{5}{6}.\,]$$

57. A club consisting of 15 married couples chooses a president and then a secretary by random selection. What is the probability that (*i*) both are men, (*ii*) one is a man and the other is a woman, (*iii*) the president is a man and the secretary is a woman, (*iv*) both are married couple? [C. H. (Econ.) '89]

[It is to be noted that the president is to be chosen first and then the secretary.

58. The nine digits $1, 2, 3, \ldots\ldots, 9$ are arranged in random order. Find the probabilily that 1, 2, 3 appear as neighbours in the order mentioned. [C. H. (Econ.) '88]

[Taking the group of 3 digits (1, 2, 3) as one digit we have $9-3+1=7$ digits in all and these can be arranged in 7 ! ways. Hence the required probability $=\dfrac{7\,!}{9\,!}=\tfrac{1}{72}$.]

59. Assume that neither A nor B has zero probability. (*i*) If A and B are mutually exclusive, will they be independent? (*ii*) If A, B are independent, will they be mutually exclusive?

[See note (*iii*), § 3.13.]

60. The probability of detecting tuberculosis in x-ray examination of a person suffering from the disease is $1-b$. The probability of diagonising a healthy person as tubercular is a. If the ratio of tubercular patients to the whole population is c, find the probability that a person is healthy if after examination he is diagonosed as tubercular. [C.H. (Math.) '94]

61. Both A and B have n coins. If they toss their coins simultaneously what is the probability that

(i) A and B will have equal number of heads ?

(ii) A will have more heads than B ?

(iii) A will have fewer heads than B ?

62. m balls are distributed among a boys and b girls. Prove that the probability that odd number of balls are distributed to boys is

$$\frac{1}{2}\frac{(b+a)^m-(b-a)^m}{(a+b)^m}.$$

Answers

5. $\frac{1}{2}, \frac{2}{15}, \frac{3}{10}, \frac{4}{5}, \frac{13}{15}$.

12. $\frac{2(n-1)!}{n!}=\frac{2}{n}$

13. (a) $1-\frac{^{22}C_{12}}{^{24}C_{12}}$, (b) $\frac{2^{12}}{^{24}C_{12}}$

14. $\frac{7}{17}$.

15. $\frac{4}{^{52}C_{13}}$.

16. $\frac{^{20}C_2\times{}^{30}C_3}{^{50}C_5}$.

17. $\frac{4}{52}\times\frac{12}{52}+\frac{8}{52}\times\frac{8}{52}+\frac{20}{52}\times\frac{12}{52}$.

20. $\frac{1}{3}\left(\frac{a_1}{a_1+b_1}+\frac{a_2}{a_2+b_2}+\frac{a_3}{a_3+b_3}\right)$.

21. $\frac{1}{N^2}\left\{\left[\frac{N}{3}\right]^2+\left(N-\left[\frac{N}{3}\right]\right)^2\right\}$,

where $[x]$ is the greatest integer not greater than x.

22. $\frac{12!}{(2!)^6\,6^{12}}$.

23. $\frac{8}{2^5}=\frac{1}{4}$.

39. $\frac{37}{56}$.

42. $\frac{1}{3}$.

57. (i) $\frac{15\times14}{30\times29}=\frac{7}{29}$, (ii) $\frac{15\times15+15\times15}{30\times29}=\frac{15}{29}$,

(iii) $\frac{15\times15}{30\times29}=\frac{15}{58}$, (iv) $\frac{2\times15}{30\times29}=\frac{1}{29}$.

60 $\frac{a(1-c)}{a+c-c(a+b)}$.

61. (i) $P(E_1)=\frac{2n!}{2^{2n}(n!)^2}$; (ii) $\frac{1}{2}\{1-P(E_1)\}$; (iii) $\frac{1}{2}\{1-P(E_1)\}$.

CHAPTER IV

COMPOUND OR JOINT EXPERIMENT

4.1. Compound or joint experiment.

Let E and E' be two random experiments and let S, S' be the respective event spaces. Now we consider the random experiment E'' which consists in performing E first and then E'. Then the event space S'' connected to E'' is given by $S''=S\times S'$. The random experiment E'' is called the compound experiment of E and E'. We observe that any outcome belonging to S'' is an ordered pair of the form (u_i, u_j') where u_i is an outcome belonging to S and u_j' is an outcome belonging to S', *i.e.*, $\{u_i\}$ is a simple event connected to E and $\{u_j'\}$ is a simple event connected to E'. So any simple event connected to the compound experiment E'' can be expressed as $\{(u_i, u_j')\}$ which is also denoted by $(\{u_i\}, \{u_j'\})$. Thus we see that if $U_i=\{u_i\}$ be a simple event connected to E and if $U_j'=\{u_j'\}$ be a simple event connected to E', then (U_i, U_j') is a simple event connected to E'' and conversely any simple event connected to E'' is of the form (U_i, U_j').

Let E be the random experiment of tossing an ordinary coin and E' be the random experiment of throwing an ordinary die. Then the corresponding event spaces S, S' are given by

$$S=\{H, T\},\ S'=\{1, 2, 3, 4, 5, 6\}.$$

So here the event space S'' connected to the compound experiment E'' (E is performed first and then E') is given by

$$S''=\{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6)\}.$$

We have defined above the compound experiment resulting from two random experiments E, E'. We can define the compound experiment $\overline{E}$ resulting from a finite number of random experiments $E_1, E_2, \ldots\ldots, E_n$ as follows :

A performance of $\overline{E}$ consists in performing $E_1, E_2, \ldots\ldots, E_n$ successively in the order mentioned. Then the event space S connected to $\overline{E}$ is given by

$$S=S_1\times S_2\times\ldots\ldots\times S_n$$

where S_i is the event space connected to E_i for $i=1, 2, \ldots\ldots, n$.

61. Both A and B have n coins. If they toss their coins simultaneously what is the probability that

(i) A and B will have equal number of heads ?

(ii) A will have more heads than B ?

(iii) A will have fewer heads than B ?

62. m balls are distributed among a boys and b girls. Prove that the probability that odd number of balls are distributed to boys is

$$\frac{1}{2}\,\frac{(b+a)^m-(b-a)^m}{(a+b)^m}.$$

Answers

5. $\frac{1}{5}, \frac{2}{15}, \frac{3}{10}, \frac{4}{5}, \frac{13}{15}$.

12. $\frac{2(n-1)\,!}{n\,!}=\frac{2}{n}$

13. (a) $1-\frac{^{22}C_{12}}{^{24}C_{12}}$, (b) $\frac{2^{12}}{^{24}C_{12}}$

14. $\frac{7}{17}$.

15. $\frac{4}{^{52}C_{13}}$.

16. $\frac{^{20}C_2 \times {}^{30}C_3}{^{50}C_5}$.

17. $\frac{4}{35}\times\frac{12}{32}+\frac{8}{35}\times\frac{8}{32}+\frac{20}{35}\times\frac{12}{32}$.

20. $\frac{1}{3}\left(\frac{a_1}{a_1+b_1}+\frac{a_2}{a_2+b_2}+\frac{a_3}{a_3+b_3}\right)$.

21. $\frac{1}{N^2}\left\{\left[\frac{N}{3}\right]^2+\left(N-\left[\frac{N}{3}\right]\right)^2\right\}$,

where $[x]$ is the greatest integer not greater than x.

22. $\frac{12\,!}{(2\,!)^6\,6^{12}}$.

23. $\frac{8}{2^5}=\frac{1}{4}$.

39. $\frac{37}{80}$.

42. $\frac{1}{3}$.

57. (i) $\frac{15\times 14}{30\times 29}=\frac{7}{29}$, (ii) $\frac{15\times 15+15\times 15}{30\times 29}=\frac{15}{29}$,

(iii) $\frac{15\times 15}{30\times 29}=\frac{15}{58}$, (iv) $\frac{2\times 15}{30\times 29}=\frac{1}{29}$.

60 $\frac{a\,(1-c)}{a+c-c(a+b)}$.

61. (i) $P(E_1)=\frac{2n\,!}{2^{2n}(n\,!)^2}$; (ii) $\frac{1}{2}\{1-P(E_1)\}$; (iii) $\frac{1}{2}\{1-P(E_1)\}$.

CHAPTER IV

COMPOUND OR JOINT EXPERIMENT

4.1. Compound or joint experiment.

Let E and E' be two random experiments and let S, S' be the respective event spaces. Now we consider the random experiment E'' which consists in performing E first and then E'. Then the event space S'' connected to E'' is given by $S''=S\times S'$. The random experiment E'' is called the compound experiment of E and E'. We observe that any outcome belonging to S'' is an ordered pair of the form (u_i, u_j') where u_i is an outcome belonging to S and u_j' is an outcome belonging to S', *i.e.*, $\{u_i\}$ is a simple event connected to E and $\{u_j'\}$ is a simple event connected to E'. So any simple event connected to the compound experiment E'' can be expressed as $\{(u_i, u_j')\}$ which is also denoted by $(\{u_i\}, \{u_j'\})$. Thus we see that if $U_i=\{u_i\}$ be a simple event connected to E and if $U_j'=\{u_j'\}$ be a simple event connected to E', then (U_i, U_j') is a simple event connected to E'' and conversely any simple event connected to E'' is of the form (U_i, U_j').

Let E be the random experiment of tossing an ordinary coin and E' be the random experiment of throwing an ordinary die. Then the corresponding event spaces S, S' are given by

$$S=\{H, T\},\ S'=\{1, 2, 3, 4, 5, 6\}.$$

So here the event space S'' connected to the compound experiment E'' (E is performed first and then E') is given by

$$S''=\{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6)\}.$$

We have defined above the compound experiment resulting from two random experiments E, E'. We can define the compound experiment $\mathbb{E}$ resulting from a finite number of random experiments $E_1, E_2, \ldots\ldots, E_n$ as follows :

A performance of $\mathbb{E}$ consists in performing $E_1, E_2, \ldots\ldots, E_n$ successively in the order mentioned. Then the event space S connected to $\mathbb{E}$ is given by

$$S=S_1\times S_2\times\ldots\ldots\times S_n$$

where S_i is the event space connected to E_i for $i=1, 2, \ldots\ldots, n$.

4.2. Independence of Random Experiments :

Let S and S' be the event spaces connected to the random experiments E and E' respectively. Let E'' be the compound experiment of performing E first and then E'. Here we assume that S and S' are at most enumerable. Two random experiments E, E' are said to be independent if for any simple event (U_i, U_j') connected to the compound experiment E'', the probability of (U_i, U_j') is given by

$$P\{(U_i, U_j')\}=P(U_i)P(U_j') \tag{4.2.1}$$

where $P(U_i)$ and $P(U_j')$ are determined respectively with respect to S and S'.

To justify the above definition of independence of two random experiments we must prove that $P(S'')=1$, where $S''=S\times S'$.

Now $P(S'')=P\left\{\sum_{i\in I}\sum_{j\in J}(U_i, U_j')\right\}$

where $S=\{u_i : i\in I\}$, $S'=\{u_j' : j\in J\}$ and $U_i=\{u_i\}$, $U_j'=\{u_j'\}$.
Here index sets I, J are at most enumerable.

$$\text{Then } P\left\{\sum_{i\in I}\sum_{j\in J}(U_i, U_j')\right\}=\sum_{i\in I}\sum_{j\in J}P\{(U_i, U_j')\}$$

$$=\sum_{i\in I}\sum_{j\in J}P(U_i)\cdot P(U_j')$$

$$=\left[\sum_{i\in I}P(U_i)\right]\left[\sum_{j\in J}P(U_j')\right]$$

$$=P(S)\,P(S')$$

$$=1\times 1=1.$$

Hence $P(S'')=1$.

So the above definition of independence of two random experiments is justified.

We can define the independence of a finite number of random experiments as follows :

The random experiments E_1, E_2, ..., E_n are said to be independent if for any simple event $(U_{i_1}, U_{i_2}, \ldots\ldots, U_{i_n})$ connected to the compound experiment E which consists in performing E_1, E_2, ..., E_n in the order mentioned,

$$P\{(U_{i_1}, U_{i_2}, \ldots\ldots, U_{i_n})\}=P(U_{i_1})P(U_{i_2})\ldots\ldots\ldots P(U_{i_n}) \tag{4.2.2}$$

where U_{i_1}, U_{i_2}, ..., U_{i_n} are simple events connected to E_1, E_2, ..., E_n respectively. The justification of this definition can be given as before.

THEOREM 4.2.1. *If A and B are two events connected to the random experiments E and E' respectively and if E and E' are independent, then*

$$P\{(A, B)\}=P(A)P(B). \tag{4.2.3}$$

Proof : Let the two random experiments E and E' be respectivly described by the event spaces S and S', where

$$S=\{u_\alpha : \alpha\in I\},\quad S'=\{u_\beta', \beta\in J\}$$

(u_α and u_β' are outcomes connected to respectively E and E'). Here we assume that the index sets I, J are at most enumerable. Then A, B can be expressed as

$$A=\sum_{\alpha\in I_1}U_\alpha,\quad B=\sum_{\beta\in J_1}U_\beta$$

where $I_1\subseteq I$, $J_1\subseteq J$ and $U_\alpha=\{u_\alpha\}$, $U_\beta'=\{u_\beta'\}$.

Now the event (A, B) connected to the compound experiment E'' can be expressed as

$$(A, B)=\left(\sum_{\alpha\in I_1}U_\alpha, \sum_{\beta\in J_1}U_\beta'\right)$$

where we note that (U_α, U_β') for all possible values of α and β are simple events connected to E''. Since E and E' are independent experiments, we have

$$P\{(U_\alpha, U_\beta')\}=P(U_\alpha)P(U_\beta').$$

Now, $P\{(A, B)\}$

$$=P\left(\left\{\sum_{\alpha\in I_1}U_\alpha, \sum_{\beta\in J_1}U_\beta'\right\}\right)=P\left\{\sum_{\alpha\in I_1}\sum_{\beta\in J_1}(U_\alpha, U_\beta')\right\}$$

$$=\sum_{\alpha\in I_1}\sum_{\beta\in J_1}P(U_\alpha, U_\beta')\quad=\sum_{\alpha\in I_1}\sum_{\beta\in J_1}P(U_\alpha)P(U_\beta')$$

$$=\left[\sum_{\alpha \varepsilon I_1} P(U_\alpha)\right]\left[\sum_{\beta \varepsilon J_1} P(U_\beta')\right]$$

$$=\left[P\left(\sum_{\alpha \varepsilon I_1} U_\alpha\right)\right]\left[P\left(\sum_{\beta \varepsilon J_1} U_\beta'\right)\right]$$

$$=P(A)P(B).$$

Hence the theorem.

4.3. Independent Trials.

Let E be a given random experiment whose event space is S. If E be repeated n times, where n is a positive integer, then the resulting experiment gives a compound experiment E_n whose event space is the cartesian product $S^n = S \times S \times \cdots \times S$ (n factors). This compound experiment E_n results from n trials of E. These n trials are said to be independent, if for any simple event $(U_{i_1}, U_{i_2}, \cdots\cdots, U_{i_n})$ connected to E_n,

$$P\{(U_{i_1}, U_{i_2}, \cdots\cdots, U_{i_n})\} = P(U_{i_1})P(U_{i_2})\cdots\cdots\cdots P(U_{i_n}) \quad (4.3.1)$$

where U_{i_r} $(r=1, 2, \cdots\cdots, n)$ is a simple event connected to E_r and where $P(U_{i_r})$ is to be determined with reference to the r-th trial of E.

Independence of n trials of E is realised in practice, if E is repeated n times under identical conditions.

4.4. Bernoulli Trials.

Let E be a random experiment where the event space S connected to E contains two distinct outcomes called 'success' and 'failure'. If E be repeated n times under identical conditions then we get n independent trials of E. These trials are called Bernoullian sequence of trials if the probability of 'success' (or 'failure') remains constant in each trial of E.

If an unbiased coin be tossed n times under identical conditions, we get Bernoulli trials where the probability of getting a 'head' (*i.e.*, success) in each trial is $\frac{1}{2}$ and that of getting a 'tail' (*i.e.*, failure) is also $\frac{1}{2}$.

4.5. THEOREM 4.5.1. (Binomial Law): *If A_i denotes the event 'exactly i successes' $(i \leq n)$ connected to the compound experiment resulting from Bernoulli trials with repetitions of the parent experiment for n times, then*

$$P(A_i) = \binom{n}{i} p^i q^{n-i} \text{ for } i = 0, 1, 2, \ldots\ldots, n \quad (4.5.1)$$

where p is the probability of 'success' in each trial and $q = 1-p$, $0 < p < 1$.

Proof: Let the random experiment E be described by the event space S, given by $S = \{s, f\}$, where 's' and 'f' are respectively the two outcomes 'success' and 'failure'. Let A_i denote the event 'exactly i successes' connected to the compound experiment E_n, resulting from n independent trials of E. Now we observe that each simple event connected to A_i is of the form $\{(s, s, f, s, f, f, \ldots f)\}$ containing i successes in any i places and failure in remaining $n-i$ places. Since the probability of success (or failure) remains constant in each trial and since trials are independent, the probability of each simple event connected to A_i is $p^i(1-p)^{n-i}$.

Further the event A_i can be expressed as the union of a number of distinct simple events of the type mentioned above. The total number of such simple events connected to A_i is equal to the number of permutations of n things of which i things are alike (success) and the remaining $(n-i)$ things are also alike (failure) but different from the former. Evidently number of such

$$\text{permutations} = \frac{n!}{i!\,(n-i)!} = \binom{n}{i}.$$

$$\therefore\ P(A_i) = p^i(1-p)^{n-i} + p^i(1-p)^{n-i} + \cdots\cdots \text{to } \binom{n}{i} \text{ times}$$

$$= \binom{n}{i} p^i (1-p)^{n-i} = \binom{n}{i} p^i q^{n-i},\quad i = 0, 1, 2, \ldots\ldots, n.$$

Note. $P(A_i) = \binom{n}{i} p^i q^{n-i}$ is the $(i+1)$-th term of the binomial expansion of $(p+q)^n$, $i = 0, 1, 2, \ldots\ldots, n$. Hence the name binomial law.

Further, the events $A_0, A_1, \ldots\ldots, A_n$ being mutually exclusive and exhaustive, we have, if S^n be the corresponding event space,

$$P(S^n) = P(A_0 + A_1 + \cdots\cdots + A_n)$$

$$=\sum_{i=0}^{n}\binom{n}{i}p^i q^{n-i}=(p+q)^n=1, \text{ since } q=1-p, \quad (4.5.2)$$

as is expected, by the concept of probability.

4.6. Poisson Approximation to Binomial Law.

THEOREM 4.6.1. *If $p=\frac{\mu}{n}$, where μ is a positive constant and $0<p<1$, n is a positive integer, then*

$$\lim_{n\to\infty}\binom{n}{i}p^i(1-p)^{n-i}=e^{-\mu}\frac{\mu^i}{i!}. \quad (4.6.1)$$

Proof: Case I. Let $i\neq o$.

We have $\binom{n}{i}p^i(1-p)^{n-i}$

$$=\frac{n!}{i!(n-i)!}\left(\frac{\mu}{n}\right)^i\left(1-\frac{\mu}{n}\right)^{n-i}, \quad \therefore \quad p=\frac{\mu}{n}$$

$$=\frac{\mu^i}{i!}\left(1-\frac{\mu}{n}\right)^n\cdot\frac{n(n-1)\ldots\ldots(n-i+1)}{n^i}\quad\frac{1}{\left(1-\frac{\mu}{n}\right)^i}$$

$$=\frac{\mu^i}{i!}\left(1-\frac{\mu}{n}\right)^n\cdot\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\ldots\left(1-\frac{i-1}{n}\right)}{\left(1-\frac{\mu}{n}\right)^i}$$

Now using the well known result, $\underset{x\to\infty}{Lt}\left(1+\frac{a}{x}\right)^{x/a}=e$, if $a\neq 0$,

we get, $\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n=\lim_{n\to\infty}\left[\left\{1+\frac{1}{\left(-\frac{n}{\mu}\right)}\right\}^{-n/\mu}\right]^{-\mu}=e^{-\mu}$.

Also $\lim_{n\to\infty}\dfrac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\ldots\left(1-\frac{i-1}{n}\right)}{\left(1-\frac{\mu}{n}\right)^i}=1.$

$$\therefore \lim_{n\to\infty}\binom{n}{i}p^i(1-p)^{n-i}=e^{-\mu}\frac{\mu^i}{i!}, \; i\neq 0.$$

Case II. If $i=0$, $\underset{n\to\infty}{Lt}\binom{n}{i}p^i(1-p)^{n-i}=\underset{n\to\infty}{Lt}\left(1-\frac{\mu}{n}\right)^n$

$$=\underset{n\to\infty}{Lt}\left\{\left(1+\frac{1}{\frac{-n}{\mu}}\right)^{-\frac{n}{\mu}}\right\}^{-\mu}$$

$$=e^{-\mu}$$

Thus it is proved that under the given conditions,

$$\underset{n\to\infty}{Lt}\binom{n}{i}p^i(1-p)^{n-i}=e^{-\mu}\frac{\mu^i}{i!}, \; i=0, 1, 2,\ldots\ldots\ldots$$

Thus if the probability of success p is small and the number of trials n is large, such that $\mu=np$ is of moderate value, we have the approximate formula

$$P(A_i)\simeq e^{-\mu}\frac{\mu^i}{i!} \quad (i=0, 1, 2,\ldots\ldots).$$

This is called *Poisson approximation* to Binomial law.

Remark: If S^∞ denotes the corresponding limiting event space, the events $A_0, A_1, A_2,\ldots\ldots$being mutually exclusive and exhaustive,

$$P(S^\infty)=P(A_0+A_1+A_2+\ldots\ldots)$$

$$=\sum_{i=0}^{\infty}e^{-\mu}\frac{\mu^i}{i!}=e^{-\mu}\sum_{i=0}^{\infty}\frac{\mu^i}{i!}=e^{-\mu}e^{\mu}=1. \quad (4.6.2)$$

4.7. Most Probable Number of Successes.

THEOREM 4.7.1. *The most probable number of successes in Bernoullian sequence of n trials is the integer (s) i_m given by the inequality*

$$(n+1)p-1 < i_m < (n+1)p \quad (4.7.1)$$

where p is the constant probability of success in each trial.

Proof: Let A_i denote the event 'i successes' connected with the compound experiment E_n.

Then $f_i=P(A_i)=\binom{n}{i}p^i(1-p)^{n-i}$

$$f_{i+1}=P(A_{i+1})=\binom{n}{i+1}p^{i+1}(1-p)^{n-i-1}$$

Now $f_{i+1} \gtreqless f_i$ according as

$$\binom{n}{i+1} p^{i+1}(1-p)^{n-i-1} \gtreqless \binom{n}{i} p^i(1-p)^{n-i}$$

i.e., according as

$$\frac{n!}{(n-i-1)!(i+1)!} p^{i+1}(1-p)^{n-i-1} \gtreqless \frac{n!}{i!(n-i)!} p^i(1-p)^{n-i}$$

i.e., according as $\frac{p}{1-p} \gtreqless \frac{i+1}{n-i}$, $i \neq n$ $\quad (\because \; 0<p<1)$

i.e., according as $np - ip \gtreqless i+1-pi-p$

i.e., according as $i \lesseqgtr (n+1)p-1$. $\quad$ (4.7.2)

Case I. Let $(n+1)\,p$ and hence $(n+1)\,p-1$ be integers. In fact $(n+1)\,p$ is a positive integer and $(n+1)\,p-1$ is a non-negative integer. Hence i can take the value $(n+1)\,p-1$. So taking

$k=(n+1)\,p-1$, $P(A_k)=P(A_{k+1})$ when $i=k=(n+1)p-1$. $\quad$ (4.7.3)

Also $k+1=(n+1)\,p>(n+1)\,p-1=k$.

$\therefore$ from (4.7.2), $P(A_{k+2})<P(A_{k+1})$.

Similarly $P(A_{k+3})<P(A_{k+2})$ and so on and finally

$$P(A_n)<P(A_{n-1}).$$

$\therefore \; P(A_k)=P(A_{k+1})>P(A_{k+2})>P(A_{k+3})>\cdots>P(A_{n-1})>P(A_n)$ $\quad$ (4.7.4)

Again $k-1<k=(n+1)\,p-1$, so that again by (4.7.2) $P(A_k)>P(A_{k-1})$. Similarly, $P(A_{k-1})>P(A_{k-2})$ and so on and finally $P(A_1)>P(A_0)$.

$\therefore \; P(A_0)<P(A_1)<P(A_2)<\cdots<P(A_{k-1})<P(A_k)=P(A_{k+1})$. (4.7.5)

Combining (4.7.4) and (4.7.5) we get

$P(A_0)<P(A_1)<P(A_2)<\cdots<P(A_{k-1})<P(A_k)$
$\quad =P(A_{k+1})>P(A_{k+2})>\cdots>P(A_{n-1})>P(A_n)$.

$\therefore \; P(A_i)$ is maximum for $i=k$ and $i=k+1$ where $k=(n+1)p-1$.

Case II. Let $(n+1)\,p$ be not an integer. Then $(n+1)\,p-1$ is also not an integer Then $(n+1)p-1<r<(n+1)p$, where $r=[(n+1)p]$, the greatest integer not greater than $(n+1)p$.

Then $r>(n+1)p-1$, which again implies, by (4.7.2),
$P(A_{r+1})<P(A_r)$.

Again $(r+1)>(n+1)\,p-1$ and so $P(A_{r+2})<P(A_{r+1})$ and so on and ultimately $P(A_n)<P(A_{n-1})$.

$\therefore \; P(A_r)>P(A_{r+1})>P(A_{r+2})>\cdots>P(A_{n-1})>P(A_n)$. $\quad$ (4.7.6)

Again $r<(n+1)p$ implies $(r-1)<(n+1)p-1$ and then by (4.7.2) $P(A_r)>P(A_{r-1})$. Similarly $P(A_{r-1})>P(A_{r-2})$ and so on and finally $P(A_1)>P(A_0)$.

$\therefore \; P(A_r)>P(A_{r-1})>P(A_{r-2})>\cdots>P(A_1)>P(A_0)$. (4.7.7)

Combining (4.7.6) and (4.7.7), we get

$P(A_0)<P(A_1)<P(A_2)<\cdots<P(A_r)>P(A_{r+1})>P(_{r+2})$
$\quad >\cdots>P(A_n)$

i.e., $P(A_i)$ is greatest when $i=r=[(n+1)p]$

i.e., $P(A_i)$ is greatest when $(n+1)p-1<i<(n+1)p$.

Combining the two cases, we find that $P(A_i)$ is maximum when i is the integer i_m, called the most probable value (*i.e.*, the *most probable number* of successes) where

$$(n+1)p-1 \leqslant i_m \leqslant (n+1)p. \quad (4.7.8)$$

4.8. Multinomial Law.

Let E be a random experiment such that the event space S connected to E has exactly m distinct outcomes, say $a_1, a_2, \ldots\ldots, a_m$. Then $S=\{a_1, a_2, \ldots\ldots, a_m\}$. The simple events connected to E are $A_1, A_2, \ldots\ldots, A_m$, where $A_i=\{a_i\}$ for $i=1, 2, \ldots\ldots, m$. Now let E be repeated n times under indentical conditions. Then we get n independent trials of E. Further we assume that the probabilities of $A_1, A_2, \ldots\ldots, A_m$ remain constant in each trial of E and let these probabilities be $p_1, p_2, \ldots\ldots, p_m$ respectively. Now let $A_{i_1 i_2 \ldots\ldots i_m}$ denote the event "A_1 occurs i_1 times, A_2 occurs i_2 times, ..., A_m occurs i_m times", where

$$i_1+i_2+\cdots\cdots+i_m=n\,;\; i_1, i_2, \ldots\ldots, i_m \in \{0, 1, 2, \ldots, n\}.$$

Then $A_{i_1 i_2 \ldots\ldots i_m}$ is an event connected to the compound experiment resulting from n independent trials of E. We observe that a particular simple event favourable to $A_{i_1 i_2 \ldots\ldots i_m}$ is

$$\{(a_1, a_1, \ldots\ldots, a_1, a_2, a_2, \ldots\ldots, a_2, \ldots\ldots, a_m, a_m, \ldots\ldots, a_m)\}$$

where a_1 occurs in the first i_1 places, a_2 occurs in the next i_2 places, ..., a_m occurs in the last i_m places. Then the probability of this simple event is $p_1^{i_1}p_2^{i_2}\ldots\ldots p_m^{i_m}$. Further we observe that the number of distinct simple events favourable to the event $A_{i_1 i_2 \ldots\ldots i_m}$ is equal to the number of permutations of n things of which i_1 are alike, i_2 are alike, ..., i_m are alike.

Now the above number of permutations is equal to

$$\frac{n!}{i_1!\; i_2!\ldots\ldots i_m!}.$$

So $P(A_{i_1 i_2 \ldots\ldots i_m}) = p_1^{i_1}p_2^{i_2}\ldots\ldots p_m^{i_m} + p_1^{i_1}p_2^{i_2}\ldots\ldots p_m^{i_m} + \ldots\ldots + p_1^{i_1}p_2^{i_2}\ldots\ldots p_m^{i_m}$,

where $p_1^{i_1}p_2^{i_2}\ldots\ldots p_m^{i_m}$ occurs $\dfrac{n!}{i_1!\; i_2!\ldots\ldots i_m!}$ times,

$$i_1+i_2+\ldots\ldots+i_m=n \quad \text{and} \quad p_1+p_2+\ldots\ldots p_m=1.$$

$$\therefore \quad P(A_{i_1 i_2 \ldots\ldots i_m}) = \frac{n!}{i_1!\; i_2!\ldots\ldots i_m!}\left(p_1^{i_1}p_2^{i_2}\ldots p_m^{i_m}\right) \tag{4.8.1}$$

where $i_1+i_2+\ldots\ldots+i_m=n$,

$p_1+p_2+\ldots\ldots+p_m=1$,

and $i_1, i_2, \ldots\ldots, i_m \in \{0,1,2,\ldots\ldots n\}$.

The above law of finding probability is known as *Multinomial law*.

Note. The above value of $P(A_{i_1 i_2 \ldots\ldots i_m})$ is the general term of the multinomial expansion of $(p_1+p_2+\ldots\ldots+p_m)^n$ and hence the formula is called the multinomial law.

Further, the events $A_{i_1 i_2 \ldots\ldots i_m}$ for all $i_1, i_2, \ldots i_m \in \{0,1,2,\ldots, n\}$ being mutually exclusive and exhaustive, we have, if S^n be the corresponding event space,

$$P(S^n) = \sum_{i_1+i_2+\ldots\ldots+i_m=n} P(A_{i_1 i_2 \ldots\ldots i_m})$$

$$= \sum_{i_1+i_2+\ldots+i_m=n} \frac{n!}{i_1!\; i_2!\ldots\ldots i_m!}\, p_1^{i_1}p_2^{i_2}\ldots\ldots p_m^{i_m}$$

$$= (p_1+p_2+\ldots\ldots+p_m)^n$$

$$= 1$$

as is expected by the concept of probability.

Deduction of Binomial Law from Multinomial Law.

Here $m=2$, $i_1=i$, $i_2=n-i$, $p_1=p$, $p_2=1-p$, since $p_1+p_2=1$. Denoting the event $A_{i_1 i_2}$ by A_i, we have by the multinomial law,

$$P(A_i)=P(A_{i_1 i_2})=\frac{n!}{i_1!\; i_2!}\, p_1^{i_1}p_2^{i_2}$$

$$=\frac{n!}{i!\,(n-i)!}\, p^i(1-p)^{n-i}.$$

Thus we get $P(A_i)=\binom{n}{i}p^i(1-p)^{n-i}$ for $i=0, 1, 2, \ldots\ldots n$.

Hence Binomial law is deduced from multinomial law.

4.9. Infinite Sequence of Bernoulli Trials.

Let $S=\{s, f\}$ be the event space corresponding to the random experiment E, where s stands for the outcome 'success' and f stands for the outcome 'failure'. Here the possible events are $L=\{s\}$, $F=\{f\}$, O and S. An infinite sequence of trials of E will be called independent if for any infinite sequence of events

$A_1, A_2, A_3, \ldots\ldots, A_n, \ldots\ldots$ connected to E,

the probability of the event $(A_1, A_2, \ldots\ldots, A_n, \ldots\ldots)$ is given by

$$P\{(A_1, A_2, \ldots\ldots, A_n, \ldots\ldots)\}=P(A_1)P(A_2)\ldots\ldots P(A_n)\ldots\ldots$$

provided the infinite product in the right hand side is convergent and where we note that each $A_i=L$ or F or S or O.

An infinite sequence of independent trials of E will be called an infinite sequence of Bernoulli trials if $P(L)$ and $P(F)$ remain constant in each trial of E.

4.10. Poisson Trials.

Let $S=\{s, f\}$ be the event space of a random experiment E where s, f denote respectively the outcomes 'success' and 'failure'. We

consider n independent trials of E where the probability of success in the k-th trial is p_k and that of failure is $1-p_k$ for $k=1, 2, \ldots\ldots, n$. If $p_1, p_2, \ldots\ldots, p_n$ are not all equal then the sequence of n independent trials of E is called a *Poisson sequence of trials.*

As an application we consider the random experiment of observing a house in a city on a given day where the possible outcomes are 'the house is not burnt on the given day' and "the house is burnt on the given day." We denote the former outcome by the letter 's' and call it a success and the latter by 'f' and call it a failure.

Let there be n houses in the city. We denote the houses by $H_1, H_2, \ldots\ldots, H_n$. Let $\frac{1}{1+i}$ be the probability that the house H_i is burnt on the given day, for $i=1, 2, \ldots\ldots, n$. Then the sequence of n independent trials which result from observing each house of the city on the given day is a poisson sequence of trials. Unlike Bernoulli trials it is difficult to obtain a general formula for finding the probability of any given number of successes in poisson sequence of trials. But we can easily find the probability of particular events connected to the compound experiment resulting from a sequence of possion trials. In the example of poisson trials mentioned above we see that the probability of the event "only the houses H_1, H_2 are burnt on the given day" is

$$\frac{1}{2}\cdot\frac{1}{3}\left(1-\frac{1}{4}\right)\left(1-\frac{1}{5}\right)\cdots\cdots\left(1-\frac{1}{n+1}\right).$$

If A_n denotes the event "n successes" in this example then

$$P(A_n)=\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{4}\right)\cdots\cdots\left(1-\frac{1}{n+1}\right).$$

4.11. Illustrative Examples.

Ex. 1. *A coin is tossed 4 times in succession. Find the probability of getting 2 heads.*

Let 'success' denote the event 'getting a head'.
Then probability of success$=p=\frac{1}{2}$,
and probability of failure$=q=\frac{1}{2}$.

$\therefore$ required probability $=\binom{4}{2}(\frac{1}{2})^2(\frac{1}{2})^2=\frac{3}{8}$.

Ex. 2. *A die is thrown 10 times in succession. Find the probability of obtaining six at least once.*

Let 'success' denote the event 'getting a six'.

Then p=probability of success$=\frac{1}{6}$

and q=probability of failure$=\frac{5}{6}$.

Let A denote the event 'at least one six'. Then the complementary event $\bar{A}$ is 'no six'.

Then $P(\bar{A})=\binom{10}{0}p^0q^{10}=(\frac{5}{6})^{10}$.

$\therefore\quad P(A)=1-(\frac{5}{6})^{10}$.

Ex. 3. *What is the probability of obtaining multiple of 3, twice in a throw of 6 dice ?*

Here p=probability of getting 'a multiple of 3' $=\frac{2}{6}=\frac{1}{3}$.

$\therefore\quad q=1-\frac{1}{3}=\frac{2}{3}$.

$\therefore$ required probability$=\binom{6}{2}(\frac{1}{3})^2(\frac{2}{3})^4=\frac{80}{243}$.

Ex. 4. *In 10 independent throws of a defective die the probability that an even number will appear 5 times is twice the probability that an even number will appear 4 times. Find the probability that an even number will not appear at all in 10 independent throws of the die.* [C.H. (Math.) '68]

Let p=probability for an even face on the die.

q=probability for an odd face on the die. $\therefore\quad p+q=1$.

Since the probability of an even number to appear five times is twice the probability that it appears 4 times, we get

$$\binom{10}{5}p^5q^5=2\times\binom{10}{4}p^4q^6$$

or, $3p=5q$, *i.e.*, $\frac{p}{5}=\frac{q}{3}=\frac{1}{8}$, $\quad\because\quad p+q=1$.

Hence probability for no even face in 10 throws

$$=\binom{10}{0}p^0q^{10}=\left(\frac{3}{8}\right)^{10}.$$

Ex. 5. *A series of tests on a certain type of electric relay have revealed that in approximately 5% of the trials, the relay fails to operate under the specified conditions. What is the probability that in ten trials made under the conditions, the relay will fail to operate one or more times?*

Let 'success' denote the event 'relay operates under specified conditions'.

Then p=probability of success=·95 and $q=1-p=·05$.

$\therefore$ probability of no failure in 10 trials

$$=\binom{10}{10}(·95)^{10}(·05)^0=(·95)^{10}.$$

$\therefore$ probability of getting one or more failures in 10 trials

$$=1-(·95)^{10}=·401.$$

Ex. 6. *At a busy street intersection, it is estimated that a joy-walker will be hit by a car with probability ·01. Assuming that individual trips form independent trials, find the probability of a joy-walker remaining unhit if he crosses the street twice per day for 50 days.* [C. H. (Math.) '84]

Let 'success' denote the event 'a joy-walker is hit by a car'.

Then p=probability of success=·01 and q=·99.

Now the joy-walker crosses the street twice per day for 50 days, *i.e.*, he crosses the road 100 times in all.

$\therefore$ required probability=probability of 'no success in 100 trials'

$$={}^{100}C_0\,(·01)^0\,(·99)^{100}=(·99)^{100}.$$

Ex. 7. *Two dice are thrown together a number of times in succession. Find the minimum number of throws to ensure that the probability of obtaining six on both the dice at least once is greater than $\frac{1}{2}$.*

When two dice are thrown, the event space S contains 36 simple events. Let the two dice be thrown n times so that the corresponding event space S^n contains 36^n simple events.

Let A denote the event 'at least one double six in n throws'. Then the complementary event $\bar{A}$ is the event 'no double six' and so $\bar{A}$ contains 35^n simple events.

$$\therefore\quad P(\bar{A})=\left(\frac{35}{36}\right)^n \text{ and so } P(A)=1-\left(\frac{35}{36}\right)^n.$$

$$\text{Now}\quad 1-\left(\frac{35}{36}\right)^n>\frac{1}{2},\quad\text{if}\quad \log\left(\frac{35}{36}\right)^n<-\log 2$$

$$i.e.,\quad \text{if } n>\frac{\log 2}{\log 36-\log 35}=24·7 \text{ nearly.}$$

Hence the least number of throws required is 25.

Ex. 8. *A missile has probability $\frac{1}{2}$ of destroying its target and probability $\frac{1}{2}$ of missing it. Assuming that the missile firings form independent trials, determine the least number of missiles that should be fired at a target in order to make the probability of destroying the target at least 0·99.* [C.H. (Math.) '82]

Let n be the number of missiles to be fired so that the probability of destroying the target is at least 0·99. Let A_r denote the event 'hitting the target at the rth trial and missing in the first $(r-1)$ trials', $r=1, 2, \ldots\ldots n$. Now,

$$P(A_r)=(1-\tfrac{1}{2})(1-\tfrac{1}{2})\ldots(1-\tfrac{1}{2})\tfrac{1}{2},\ \text{where } (1-\tfrac{1}{2}) \text{ occurs } (r-1) \text{ times,}$$

$$=\frac{1}{2^r},\ \text{for } r=1, 2, \ldots\ldots n.$$

Let B denote the event 'the target is destroyed in n trials'.

Then $B=A_1+A_2+\cdots\cdots+A_n$, where $A_1, A_2, \ldots\ldots, A_n$ are pairwise mutually exclusive.

$$\therefore\quad P(B)=\sum_{r=1}^{n}P(A_r)=\frac{1}{2}+\frac{1}{2^2}+\cdots\cdots+\frac{1}{2^n}=1-\frac{1}{2^n}.$$

Then $P(B)\geqslant ·99$ if $1-\dfrac{1}{2^n}\geqslant ·99$ *i.e.*, if $(·5)^n\leqslant ·01$;

$$i.e.,\quad \text{if } n\geqslant\frac{2}{\log 2}=6·4 \text{ nearly.}$$

$\therefore$ required least number of missiles that should be fired $=7$.

Ex. 9. *Suppose that probability of a new-born baby a boy is ·45. In a family of 8 children, calculate the probability that (i) there are 3 or 4 boys (ii) number of boys lies in the interval* [3, 7].

We consider successive births of babies in a family as independent Bernoulli trials. If 'success' means the event 'a new-born baby a boy', we have 8 Bernoulli trials with p=probability of success=·45 and $q=·55$.

(*i*) If A be the event 'there are 3 or 4 boys', then

$$P(A)={}^8C_3\,p^3q^5+{}^8C_4\,p^4q^4=\cdot519.$$

(*ii*) $A_r \equiv$ 'r boys out of 8 children', then required probability

$$=P\left(\sum_{r=3}^{7} A_r\right)=\sum_{r=3}^{7} P(A_r)$$

$$=1-P(A_0)-P(A_1)-P(A_2)-P(A_8)$$

$$=1-(\cdot55)^8-\binom{8}{1}(\cdot45)(\cdot55)^7-\binom{8}{2}(\cdot45)^2(\cdot55)^6-\binom{8}{8}(\cdot45)^8(\cdot55)^0$$

$$=0\cdot778$$

Ex. 10. *Suppose that the probability of an item being defective, in a mass production process of that item is 0·01. If 20 items are selected at random, then what is the probability that exactly 2 will be defective?*

By Binomial law, probability of getting exactly 2 defectives

$$=\binom{20}{2}(\cdot01)^2(\cdot99)^{18}=\cdot0158.$$

[Again by poisson approximation to Binomial law, where

$$\mu=np=\cdot2,$$

probability of getting exactly two defectives

$$=e^{-\mu}\frac{\mu^2}{2!}=\frac{(\cdot2)^2}{2}e^{-\cdot2}=\cdot0164.$$

The error in the approximation $=\cdot0164-\cdot0158=\cdot0006.$]

Ex. 11. *A system consists of 1000 connected components, where each component may fail independently of the others. If the probability that a component fails in one month is 10^{-3}, then find the probability that the system will function (i.e., no component will fail) throughout a month.*

If 'success' means the event 'a component fails in one month', then probability of success being small and $n=1000$ being large, we apply poisson approximation to binomial law. The probability of success $=p=10^{-3}$ and $\mu=np=1000\times10^{-3}=1$.

$\therefore$ required probability $=e^{-\mu}\dfrac{\mu^0}{0!}=e^{-1}=\cdot368.$

Ex. 12. *The probability of success in any trial of a given random experiment is $\frac{1}{3}$. If p_n be the probability of getting no two consecutive successes in n trials of the given experiment, then prove that* $p_n=\frac{2}{3}p_{n-1}+\frac{2}{9}p_{n-2}$.

Let A_n denote the event 'no two consecutive successes in n trials.' Then A_n can be expressed as $A_n=B_1+B_2$ where B_1 and B_2 are two mutually exclusive events defined by

$B_1\equiv$'failure in the nth trial and no two consecutive successes in the first $(n-1)$ trials'

and $B_2\equiv$'success in the nth trial, failure in $(n-1)$th trial and no two consecutive successes in the first $(n-2)$ trials'.

Then $P(A_1)=P(B_1)+P(B_2)$, where

$P(B_1)=$probability of failure in the nth trial $\times$ probability of no two consecutive successes in the first $(n-1)$ trials

$$=\tfrac{2}{3}\,p_{n-1},$$

and $P(B_2)=$probability of success in the nth trial $\times$ probability of failure in the $(n-1)$th trial $\times$ probability of no two consecutive successes in the first $(n-2)$ trials

$$=\tfrac{1}{3}\times\tfrac{2}{3}\times p_{n-2}=\tfrac{2}{9}\,p_{n-2}.$$

$$\therefore\quad p_n=\tfrac{2}{3}\,p_{n-1}+\tfrac{2}{9}\,p_{n-2}.$$

Ex. 13. *A class has only three students A, B, C who attend the class independently, the probability of their attendance on any day being $\frac{1}{3}$, $\frac{2}{3}$, $\frac{3}{4}$ respectively. Find the probability that the total number of attendances in two consecutive days is exactly three.*

[*C.H.* (*Math.*) '80]

Let E_1 be the random experiment of observing if A attends the class in any day. Let s denote the outcome 'A attends the class in a day' and call it a success and f denote the outcome 'A does not attend the class in the day' and call it a failure. Then observing the attendance of A on two consecutive days we get Bernoulli trials with $n=2$ and $p=\frac{1}{3}$. Similarly we define the random experiments E_2, E_3 and the corresponding Bernoulli trials with $n=2$, $p=\frac{2}{3}$ and $n=2$, $p=\frac{3}{4}$ considering the attendances of B and C respectively. We denote the three compound experiments resulting from above Bernoulli trials respectively by

$$E_1^{(2)},\quad E_2^{(2)},\quad E_3^{(2)}.$$

Now the required event can be decomposed into the following events :

I : $E_1^{(2)}$: 2f II : $E_1^{(2)}$: 2f III : $E_1^{(2)}$: 2s IV : $E_1^{(2)}$: 2s

$E_2^{(2)}$: 1s, 1f $E_2^{(2)}$: 2s $E_2^{(2)}$: 1s, 1f $E_2^{(2)}$: 2f

$E_3^{(2)}$: 2s $E_3^{(2)}$: 1s, 1f $E_3^{(2)}$: 2f $E_3^{(2)}$: 1s, 1f

V : $E_1^{(2)}$: 1s, 1f VI : $E_1^{(2)}$: 1s, 1f VII : $E_1^{(2)}$: 1s, 1f

$E_2^{(2)}$: 2s $E_2^{(2)}$: 2f $E_2^{(2)}$: 1s, 1f

$E_3^{(2)}$: 2f $E_3^{(2)}$: 2s $E_3^{(2)}$: 1s, 1f.

$\therefore$ the required probability$=(\frac{1}{2})^2\binom{2}{1}(\frac{2}{3})(\frac{1}{3})(\frac{3}{4})^2+(\frac{1}{2})^2(\frac{2}{3})^2\binom{2}{1}(\frac{3}{4})(\frac{1}{4})$

$+(\frac{1}{2})^2\binom{2}{1}(\frac{2}{3})(\frac{1}{3})(\frac{1}{4})^2+(\frac{1}{2})^2(\frac{1}{3})^2(\frac{1}{3})^2\binom{2}{1}(\frac{3}{4})(\frac{1}{4})+\binom{2}{1}(\frac{1}{2})(\frac{1}{2})(\frac{2}{3})^2(\frac{1}{4})^2$

$+\binom{2}{1}(\frac{1}{2})(\frac{1}{2})(\frac{1}{3})^2(\frac{3}{4})^2+\binom{2}{1}(\frac{1}{2})(\frac{1}{2})\binom{2}{1}(\frac{2}{3})(\frac{1}{3})\binom{2}{1}(\frac{3}{4})(\frac{1}{4})$.

$=\frac{1}{16}+\frac{1}{24}+\frac{1}{144}+\frac{1}{96}+\frac{1}{72}+\frac{1}{32}+\frac{1}{12}=\frac{1}{4}$.

Ex. 14. *If r numbers are selected at random from ten numbers 1, 2, 3,……, 10 repetitions being allowed, then show that the probability that the product of the numbers is of the form $3n-1$ or $3n+1$ (n being a non-negative integer) is $(\frac{7}{10})^r$.*

Any integer is always of the form $3n$ or $3n-1$ or $3n+1$, n being a non-negative integer and the product of the r numbers selected is of the form $3n$, if at least one number selected is 3 or 6 or 9.

Let us consider the random experiment of selecting a number from 1, 2, 3,……, 10 and let 'success' means 'numbers selected is a multiple of 3'.

Then p = probability of success $=\frac{3}{10}$,

q = probability of failure $=\frac{7}{10}$.

Now in r repetitions of the above experiment, the required event is the event "no success" and so the required probability

$$={}^rC_0 p^0 q^r=(\tfrac{7}{10})^r.$$

Ex. 15. *Three coins having probabilities of head $\frac{1}{2}$, $\frac{2}{5}$, $\frac{3}{7}$ respectively are thrown. Find the probability of obtaining exactly one head and two heads.*

Here $p_1=\frac{1}{2}$, $p_2=\frac{2}{5}$, $p_3=\frac{3}{7}$, where p_i denotes the probability of getting 'head' with the ith coin for $i=1, 2, 3$.

$\therefore$ p (exactly one head) $=p_1q_2q_3+q_1p_2q_3+q_1q_2p_3$

$$=\tfrac{1}{2}\times\tfrac{3}{5}\times\tfrac{4}{7}+\tfrac{1}{2}\times\tfrac{2}{5}\times\tfrac{4}{7}+\tfrac{1}{2}\times\tfrac{3}{5}\times\tfrac{3}{7}=\tfrac{29}{70},$$

p (exactly two heads) $=p_1p_2q_3+p_1q_2p_3+q_1p_2p_3$

$$=\tfrac{1}{2}\times\tfrac{2}{5}\times\tfrac{4}{7}+\tfrac{1}{2}\times\tfrac{3}{5}\times\tfrac{3}{7}+\tfrac{1}{2}\times\tfrac{2}{5}\times\tfrac{3}{7}=\tfrac{23}{70}.$$

Ex. 16. *Find the most probable number of times the event 'multiple of three' will occur when a die is thrown 100 times.*

Following theorem 4.7.1 if i_m be the most probable number, then $(n+1)p-1\le i_m\le(n+1)p$, where $n=100$,

p = probability of getting a multiple of three $=\frac{2}{6}=\frac{1}{3}$.

$\therefore$ $(n+1)p-1=\frac{101}{3}-1=\frac{98}{3}$ and $(n+1)p=\frac{101}{3}$

$\therefore$ $i_m=33$.

Ex. 17. *Find the probability that in 8 throws of a die, the number 1, 3, 5 turn up 2, 3, 3 times respectively.*

From multinomial law, the required probability

$$=\frac{8\,!}{2\,!\;3\,!\;3\,!}(\tfrac{1}{6})^2(\tfrac{1}{6})^3(\tfrac{1}{6})^3,$$

since the probabilities of getting 1,3,5 from a single throw are all equal to $\frac{1}{6}$.

Ex. 18. (**Drawing with replacement**) : *An urn contains N balls, of which N_1 are red and N_2 are black. If n balls are drawn successively from the urn and are replaced each time, then find the probability that r drawings will yield red balls.*

$(N=N_1+N_2,\ 0<n\le N,\ 0\le r\le \min.(n, N_1)$.

Let A_r denote the event 'r red balls drawn' in n trials.

Now p=probability of drawing a red ball$=\dfrac{N_1}{N}$ and

q=probability of drawing a black ball$=\dfrac{N_2}{N}$.

The balls being replaced after each trial, the trials are independent and

$$P(A_r)=\binom{n}{r}p^r q^{n-r}=\binom{n}{r}\left(\frac{N_1}{N}\right)^r\left(\frac{N_2}{N}\right)^{n-r}.$$

Ex. 19. *An urn contains 1 white and 99 black balls. If 1000 drawings are made with replacements, then what is the probability that 10 drawings will yield white balls ?*

Here $p=\frac{1}{100}$, $n=1000$ and $\mu=np=10$. Then p is small, n is large while μ is of moderate value. So we apply poisson approximation.

Required probability $\simeq e^{-10}\dfrac{10^{10}}{10\,!}$.

Ex. 20. (Drawing without replacement). *An urn contains N balls of which N_1 are red and N_2 are black. If n balls are drawn successively from the urn without replacement, then find the probability that r drawings will yield red balls. Hence show that binomial law can be obtained as a limiting case.*

$$[N=N_1+N_2,\ 0<n\le N,\ 0\le r\le \min(n, N_1)].$$

If n balls are drawn without replacement, it amounts to drawing n balls simultaneously.

$\therefore$ probability of drawing r red balls $=\dfrac{\binom{N_1}{r}\binom{N_2}{n-r}}{\binom{N}{n}}$.

Now
$$\frac{\binom{N_1}{r}\binom{N_2}{n-r}}{\binom{N}{n}}=\frac{N_1(N_1-1)\dots(N_1-r+1)}{r\,!}$$
$$\times\frac{N_2(N_2-1)\dots(N_2-n+r+1)}{(n-r)\,!}$$
$$\times\frac{n\,!}{N(N-1)\dots(N-n+1)}$$
$$=\frac{n\,!}{r\,!\,(n-r)\,!}\times\frac{N_1(N_1-1)\dots(N_1-r+1)}{N_1^r}\times N_1^r$$
$$\times\frac{N_2(N_2-1)\dots(N_2-n+r+1)}{N_2^{n-r}}\times N_2^{n-r}$$
$$\times\frac{N^n}{N(N-1)\dots(N-n+1)}\times\frac{1}{N^n}$$

$$=\binom{n}{r}\frac{N_1^r N_2^{n-r}}{N^n}\times\left(1-\frac{1}{N_1}\right)\left(1-\frac{2}{N_1}\right)\dots\left(1-\frac{r-1}{N_1}\right)$$
$$\times\left(1-\frac{1}{N_2}\right)\left(1-\frac{2}{N_2}\right)\dots\left(1-\frac{n-r-1}{N_2}\right)$$
$$\times\frac{1}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{n-1}{N}\right)}$$
$$=\binom{n}{r}\left(\frac{N_1}{N}\right)^r\left(\frac{N_2}{N}\right)^{n-r}\times\left(1-\frac{1}{N_1}\right)\left(1-\frac{2}{N_1}\right)\dots\left(1-\frac{r-1}{N_1}\right)$$
$$\times\left(1-\frac{1}{N_2}\right)\left(1-\frac{2}{N_2}\right)\dots\left(1-\frac{n-r-1}{N_2}\right)$$
$$\times\frac{1}{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{n-1}{N}\right)}$$

$\therefore\ \underset{N\to\infty}{Lt}\dfrac{\binom{N_1}{r}\binom{N_2}{n-r}}{\binom{N}{n}}=\binom{n}{r}p^r q^{n-r}$, if $p=\dfrac{N_1}{N}$, $q=\dfrac{N_2}{N}$ are fixed and so $N_1\to\infty$, $N_2\to\infty$ as $N\to\infty$, which is the probability of r successes in n trials. Thus binomial law can be obtained as a limiting case for drawing without replacement.

Ex. 21. *If a die is thrown n times, then find the probability that*
(i) *the greatest,*
(ii) *the least number obtained will have a given value r.*

(i) Let X denote the event 'greatest value is r'.

Also let A_i denote the event "each value is less than or equal to i".

Then we can write $A_r=A_{r-1}+X$, where A_{r-1}, X are mutually exclusive events.

So $P(A_r)=P(A_{r-1})+P(X)$

or, $P(X)=P(A_r)-P(A_{r-1})$

Now $P(A_r)=\dfrac{r^n}{6^n}$ since A_r happens if and only if the value in each throw belongs to the set $\{1, 2, \dots\dots, r\}$.

Similarly $P(A_{r-1})=\dfrac{(r-1)^n}{6^n}$.

Hence $P(X)=\dfrac{r^n-(r-1)^n}{6^n}$.

(*ii*) Let Y denote the event "least value is r". If B_i denote the event "each value is greater than or equal to i" then we can write $B_r=B_{r+1}+Y$, where Y, B_{r+1} are mutually exclusive.

Hence $P(B_r)=P(B_{r+1})+P(Y)$.

Now $P(B_r)=\dfrac{(6-r+1)^n}{6^n}$ and $P(B_{r+1})=\dfrac{(6-r)^n}{6^n}$.

So $P(Y)=\dfrac{(6-r+1)^n}{6^n}-\dfrac{(6-r)^n}{6^n}$.

$$=\frac{(7-r)^n}{6^n}-\frac{(6-r)^n}{6^n}.$$

Ex. 22. *In a Bernoullian sequence of n trials with probability of success p, find the probability that the ith success occurs at the nth trial.*

Let A denote the event 'ith success occurs at the nth trial'.

Let A_1 denote the event '$i-1$ successes in the first $n-1$ trials' and B denote the event 'success in the nth trial.' Then A happens if and only if A_1 and B both happen. So

$P(A)=P(A_1\ B)=P(A_1)\ P(B)$, since the trials are independent.

Now $P(A_1)={}^{n-1}C_{i-1}\ p^{i-1}\ (1-p)^{n-i}$ and $P(B)=p$.

So the required probability is

$${}^{n-1}C_{i-1}\ p^{i-1}(1-p)^{n-i}.p={}^{n-1}C_{i-1}\ p^{i}(1-p)^{n-i}.$$

Examples IV

1. Find the probability of (*i*) getting 3 sixes, (*ii*) getting at least 3 sixes in 5 throws of a single die.

2. A coin is tossed 8 times in succession. What is the probability of obtaining (*i*) 5 heads, (*ii*) 3 heads and 5 tails, (*iii*) at least 2 heads?

3. A and B play a game in which A's chance of winning is $\frac{2}{3}$. In a series of 8 games, what is the chance that A will win at least 6 games?

4. If there are on the average 5 per cent colour-blind, then what is the chance of having at least 2 colour-blinds in the population of 100 people?

5. If there is on the average 1 left-hander in a population of 100, then what is the probability of getting 3 left-handers in a population of 150?

6. Show that the probability of obtaining no head in an infinite sequence of independent tosses of a coin is zero.

[*C.H.* (*Math.*) '91]

[*Hints*: p_n = probability of getting no head in n independent tosses $=\left(1-\dfrac{1}{2}\right)\left(1-\dfrac{1}{2}\right)\ldots\ldots n$ times $=\dfrac{1}{2^n}$.

Required probability $=\underset{n\to\infty}{Lt}\ \dfrac{1}{2^n}=0.$]

7. If the probability of hitting a target is $\frac{1}{5}$ and 10 shots are fired independently, then what is the probability of the target being shot at least twice?

8. The probability of hitting a target is 0·001 for each shot. Find the probability of hitting the target with two or more bullets if the number of shots is 5,000.

[*Hints*: Applying Poisson approximation to binomial law, here $n=5000$, $\mu=5000\times{}\cdot001=5$.

Required probability $=1-\{P(A_0)+P(A_1)\}$, where A_i is the event 'i successes in 5,000 trials'

$$=1-e^{-5}-5e^{-5}=1-6e^{-5}\simeq\cdot9596.]$$

9. The probability of a man hitting a target is $\frac{1}{4}$. How many times he should fire so that the probability of his hitting the target at least once is greater than $\frac{2}{3}$?

Hints: If A_i be the event 'target is hit i times from n firings', then the probability of hitting the target at least once

$$=P(A_1+A_2+\ldots\ldots+A_n)=1-P(A_0)=1-{}^nC_0(\tfrac{1}{4})^0(\tfrac{3}{4})^n.$$

Now, $1-(\frac{3}{4})^n > \frac{2}{3}$

if $(\frac{3}{4})^n < \frac{1}{3}$

i.e., if $n > \dfrac{\log 3}{\log 4-\log 3}=3{\cdot}8.$

$\therefore\ n=4.$

10. The probability of a man hitting a target is $\frac{1}{3}$. How many times must he fire so that the probability of hitting the target at least once is more that 90% ?

11. If the chances of a child being male or being female are equal, then find the probability that among 4 children of a family two are boys and two are girls.

12. A lottery organisation has announced 100 prizes out of 10,000 tickets. Find the minimum number of tickets a person should purchase so that the probability of his getting at least one prize is greater than $\frac{3}{4}$.

[*Hints* : Purchasing tickets one after another form Bernoulli trials, probability of success $p=\frac{1}{100}$.

Probability of at least one success from n tickets $=1-P(A_0)$ (where A_i is the event 'i successes in n trials') $=1-(\frac{99}{100})^n$.

$\therefore\ 1-(\frac{99}{100})^n > \frac{3}{4}$ or, $(\frac{99}{100})^n < \frac{1}{4}$

or, $n > \dfrac{\log 4}{2-\log 99} \simeq 137{\cdot}935$

$\therefore\ n=138$.]

13. In a screw manufacturing factory, the probability that a screw is defective is ·02. 100 screws are taken for inspection from a box. Find the probability that (*i*) there is no defective screw; (*ii*) at most two defective screws; (*iii*) exactly four defective screws.

[*Hints* : The process of inspecting screws from a box one after another form Bernoulli trials. Let the event 'a screw is defective' be called a success. Then $p=$ probability of success $=\cdot 02$. Using Poisson approximation to binomial law, here $n=100$, $\mu=100\times\cdot 02=2$,

(*i*) required probability $=e^{-\mu}\,\dfrac{\mu^0}{0\,!}=e^{-2}=\cdot 135.$

(*ii*) if A_r denotes the event 'r successes in 100 trials', the required probability $=P(A_0+A_1+A_2)$

$=P(A_0)+P(A_1)+P(A_2)$, the events A_0, A_1, A_2 being mutually exclusive.

$=e^{-\mu}\left(\dfrac{\mu^0}{0\,!}+\dfrac{\mu}{1\,!}+\dfrac{\mu^2}{2\,!}\right)$

$=5e^{-2}=\cdot 676,$

(*iii*) required probability $=e^{-\mu}\cdot\dfrac{\mu^4}{4\,!}=e^{-2}\cdot\frac{2}{3}=\cdot 09.$]

14. *A* coin with constant probability p of showing head in any tossing is tossed $2n$ times. Find the most probable number of heads. [*C.H.* (*Math.*) '80, '88]

[*Hints* : If i_m be the most probable number, then

$(2n+1)p-1 \le i_m \le (2n+1)p$ by Theorem 4.8.1

If $(2n+1)\,p$ is a positive integer, then $i_m=(2n+1)p$, $(2n+1)p-1$.

If $(2n+1)\,p$ is not a positive integer, then $i_m=$ integral part of $(2n+1)p$.

In particular, if the coin is a fair one, $p=\frac{1}{2}$, $i_m=n$.]

15. (*a*) If a day is dry, then the conditional probability that the following day is dry is p ; if a day is wet, the conditional probability that the following day is dry is p' ; if u_n is the probability that the nth day will be dry, then prove that $u_n-(p-p')\,u_{n-1}-p'=0,\ n \geqslant 2.$

(*b*) If the first day is sure to be dry, $p=\frac{3}{4}$, $p'=\frac{1}{4}$, then find u_n. [*C.H.* (*Math.*) '79]

[*Hints* : Let A_1, A_2 and X denote the events '$(n-1)$th day dry', '$(n-1)$th day wet' and 'nth day dry' respectively.

Then $P(X \mid A_1)=p$, $P(X \mid A_2)=p'$, $P(X)=u_n$, $P(A_1)=u_{n-1}$ and $P(A_2)=1-u_{n-1}$.

(*a*) Now $X=XA_1+XA_2$ and the two events on the right hand side are mutually exclusive.

$\therefore\ P(X)=P(XA_1)+P(XA_2)=P(A_1)P(X \mid A_1)+P(A_2)P(X \mid A_2)$

i.e., $u_n=pu_{n-1}+p'(1-u_{n-1})$

or, $u_n-(p-p')u_{n-1}-p'=0,\ n \geqslant 2.$

(b) $u_n - (\frac{3}{4} - \frac{1}{4})u_{n-1} - \frac{1}{4} = 0$

or, $2^2 u_n - 2u_{n-1} - 1 = 0$

or, $2^{n+1} u_n - 2^n u_{n-1} - 2^{n-1}(2-1) = 0$

or, $2^n(2u_n - 1) = 2^{n-1}(2u_{n-1} - 1) = \cdots\cdots = 2(2u_1 - 1) = 2$,

as the first day is sure to be dry, $u_1 = 1$.

$\therefore \quad u_n = \frac{1}{2^n} + \frac{1}{2}$.]

16. A player tosses a coin and is to score one point for every head turned up and two for every tail. He is to play until his score reaches or passes n. If p_n is his chance of attaining exactly n, then show that

$$p_n = \tfrac{1}{2}(p_{n-1} + p_{n-2}).$$

Hence find p_n and its limit as $n \to \infty$. [*C.H.* (*Math.*) '63, '67]

[*Hints*: Let B_1, B_2 and A_n be the events 'head turns up', 'tail turns up' in any trial and 'score is exactly n' respectively.

Now a player can score exactly n in two ways:

(*i*) scoring $n-2$ at a certain stage and then getting a tail in the next trial,

or, (*ii*) getting a head in the next trial when he had just scored $n-1$.

$\therefore \quad A_n = A_{n-1}B_1 + A_{n-2}B_2$ and the two events on the right hand side being mutually exclusive and the trials being independent,

$$p_n = \tfrac{1}{2}(p_{n-1} + p_{n-2}).$$

$\therefore \quad p_n + \frac{1}{2}p_{n-1} = p_{n-1} + \frac{1}{2}p_{n-2} = \cdots\cdots = p_2 + \frac{1}{2}p_1$.

Now p_2 = probability of scoring 2

$= P(B_2 + B) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$,

where B denotes 'head in the first trial and head in the second trial'

and p_1 = probability of scoring $1 = \frac{1}{2}$.

$\therefore \quad p_n + \frac{1}{2}p_{n-1} = \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{2} = 1$.

To find p_n we write, $p_n - k = -\frac{1}{2}(p_{n-1} - k)$

i.e., $p_n + \frac{1}{2}p_{n-1} = \frac{3}{2}k$.

$\therefore \quad \frac{3}{2}k = 1 \quad$ i.e., $k = \frac{2}{3}$.

$\therefore \quad p_n - \frac{2}{3} = -\frac{1}{2}(p_{n-1} - \frac{2}{3})$.

Replacing n by $n-1, n-2, \ldots\ldots, 2$ successively, we get

$p_{n-1} - \frac{2}{3} = -\frac{1}{2}(p_{n-2} - \frac{2}{3})$

$p_{n-2} - \frac{2}{3} = -\frac{1}{2}(p_{n-3} - \frac{2}{3})$.

... ...

$p_2 - \frac{2}{3} = -\frac{1}{2}(p_1 - \frac{2}{3})$

From the above relations, we get

$p_n - \frac{2}{3} = (-\frac{1}{2})^{n-1}(p_1 - \frac{2}{3}) = (-\frac{1}{2})^n \cdot \frac{1}{3}$

i.e., $p_n = \frac{1}{3}\left\{2 + (-1)^n \cdot \frac{1}{2^n}\right\}$

$\therefore \quad \underset{n \to \infty}{Lt} \; p_n = \frac{2}{3}$.]

17. An urn contains a white and b black balls, and a series of drawings of one ball made at a time, the ball removed being returned to the urn immediately after the next drawing is made. If p_n denotes the probability that the nth ball drawn is black, then show that

$$p_n = \frac{b - p_{n-1}}{a+b-1}.$$

[*C.H.* (*Math.*) '70]

[*Hints*: When the nth ball is drawn, the ball drawn in $(n-1)$th drawing is still not returned.

Hence the appearance of a black ball in the nth trial is followed by the event '$(n-1)$th ball is black or white.'

$$p_n = p_{n-1}\frac{b-1}{a+b-1} + (1 - p_{n-1})\frac{b}{a+b-1} = \frac{b - p_{n-1}}{a+b-1}.]$$

18. A die is thrown 10 times in succession. Find the probability of the occurrence of six 4 times, five twice and all other faces once each.

19. Prove that the probability of nth success at the time of $(n+k)$th trial of Bernoulli sequence is $\binom{n+k-1}{k}p^n q^k$, where p is the probability of success in each trial and $q = 1-p$.

[*Hints*: Let A, B, C denote the events 'nth success at the $(n+k)$th trial', 'k failures in $(n+k-1)$ trials' and 'success at the $(n+k)$th trial' respectively.

Then $A = BC$.

$\therefore$ $P(A) = P(B)P(C)$, the trials being independent.

$\therefore$ $P(A) = \binom{n+k-1}{k} q^k (1-q)^{n-1} \cdot p = \binom{n+k-1}{k} p^n q^k$.]

20. **Banach Match Box Problem**: A mathematician always carries two match boxes, each containing n matches. Whenever he needs, he chooses a box at random and draws a match from it. Find the probability that when the first box is found to be empty for the first time, the second box will contain exactly i matches. [C.H. (Math.) '66]

[*Hints*: When the first box was found empty for the first time, it was chosen $(n+1)$ times so that n matches were drawn already and at the time of $(n+1)$th selection it was found empty. Also since the second box contains exactly i matches, it was selected $(n-i)$times. Hence if we consider the random experiment of choosing a box at random we have $(n+1)+(n-i) = 2n-i+1$ trials, where we call the event 'selecting the first box' a success.

$\therefore$ required probability
= probability of n successes in $2n-i$ trials and success at the $(2n-i+1)$th trial

$= \binom{2n-i}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^{n-i}\left(\frac{1}{2}\right)$, the trials being independent

$= \binom{2n-i}{n}\left(\frac{1}{2}\right)^{2n-i+1}$.]

21. In a sequence of Bernoulli trials, with probability p for success, what is the probability of 'a' successes before 'b' failures?

22. From an urn containing n tickets numbered $1, 2, \ldots\ldots, n$; k tickets are drawn at a time and replaced before the next drawing. Find the probability that in r such drawings, ticket numbers $1, 2, \ldots\ldots, r$ do not appear in the 1st, 2nd,......, rth drawings respectively.

[*Hints*: Let E_i denote the event 'ith ticket does not appear in the ith drawing of k tickets'; where $i = 1, 2, \ldots\ldots, r$.

Then $P(E_i) = \binom{n-1}{k} \Big/ \binom{n}{k} = \frac{n-k}{n}$.

Now the required event is $(E_1, E_2, \ldots\ldots, E_r)$ of the corresponding compound experiment of r independent trials of the above experiment.

$\therefore$ $P\{(E_1, E_2, \ldots\ldots, E_r)\} = P(E_1)P(E_2)\ldots\ldots P(E_r)$,
the trials being independent

$= \left(\frac{n-k}{n}\right)^r$.]

23. A and B play a game which must be either won or lost. If the probability that A wins a game is P, then find the probability that A wins m games before B wins n games $(m, n \geqslant 1)$.

[*Hints*: Let E be the event 'A wins m games before B wins n games'. We consider $m+n-1$ Bernoulli trials with the event 'A wins' as success. Then E happens if and only if in $(m+n-1)$ trials we have at least m suecesses.

The required probability $= \sum\limits_{i=m}^{m+n-1} \binom{m+n-1}{i} p^i (1-p)^{m+n-i-1}$]

24. An urn contains n tickets numbered 1 to n, from which a ticket is drawn and replaced r times. What is the probability that the greatest number drawn is i?

[*Hints*: We consider the random experiment of drawing r tickets one after another with replacement as Bernoulli trials and call the event 'number drawn is less than or equal to i' as success.

Then p = probability of success $= \frac{i}{n}$.

Let X_i be the event 'the greatest number drawn is less than or equal to i'. Then X_i happens when each of the r drawings yields a ticket whose number is less than or equal to i.

$\therefore$ $P(X_i) = \left(\frac{i}{n}\right)^r$.

$\therefore$ required probability $= P(X_i) - P(X_{i-1})$

$= \frac{i^r - (i-1)^r}{n^r}$.]

25. Find the most probable number of heads in 10 throws of a biased coin, the probability of getting a head in a single throw is $\frac{1}{3}$.

Answers

1. (*i*) $\binom{5}{3}(\frac{1}{6})^3(\frac{5}{6})^2=\frac{125}{3888}$;

(*ii*) $\binom{5}{3}(\frac{1}{6})^3(\frac{5}{6})^2+\binom{5}{4}(\frac{1}{6})^4(\frac{5}{6})+\binom{5}{5}(\frac{1}{6})^5=\frac{23}{648}$.

2. (*i*) $\binom{8}{3}(\frac{1}{2})^5(\frac{1}{2})^3=\frac{7}{32}$; (*ii*) $\binom{8}{5}(\frac{1}{2})^5(\frac{1}{2})^3=\frac{7}{32}$;

(*iii*) $1-\{{}^8C_0(\frac{1}{2})^8+{}^8C_1(\frac{1}{2})(\frac{1}{2})^7\}=1-\frac{9}{2^8}=\frac{247}{256}$.

3. If A_i denotes the event 'A wins i games' the required probability$=P(A_6)+P(A_7)+P(A_8)=\sum_{r=6}^{8}\binom{8}{r}\left(\frac{2}{3}\right)^r\left(\frac{1}{3}\right)^{8-r}$.

4. The probability of any member of the population being colour-blind $=\frac{5}{100}=\frac{1}{20}$.

$\therefore$ the required probability$=1-\binom{100}{0}(\frac{19}{20})^{100}-\binom{100}{1}(\frac{1}{20})(\frac{19}{20})^{99}$.

5. $\binom{150}{3}\left(\frac{1}{100}\right)^3\left(\frac{99}{100}\right)^{147}$ 7. $\sum_{r=2}^{10}\binom{10}{r}(\cdot 2)^r(\cdot 8)^{10-r}$

10. 6. 11. ${}^4C_2(\frac{1}{2})^2(\frac{1}{2})^2=\frac{3}{8}$.

18. $\frac{10!}{4!\,2!\,(1!)^4}\left(\frac{1}{6}\right)^{10}$ by multinomial law.

21. $\sum_{i=a}^{a+b-1}\binom{a+b-1}{i}p^iq^{a+b-1-i}$. 25. 3.

CHAPTER V

PROBABILITY DISTRIBUTIONS

5·1. Random Variables:

We know that the results of a random experiment are objects which are not necessarily real numbers and so it will not be easy to develop a mathematical theory dealing with these objects. Now it is possible to establish a correspondence between outcomes of a random experiment and a given set of real numbers by means of a mapping defined on the event space of the given random experiment as domain and by such correspondence it will be convenient to use algebra and analysis of real numbers in the development of the mathematical theory of probability. With this motivation we introduce the concept of a random variable connected to the event space of any given random experiment.

Definition of a random variable:

Let E be a random experiment and S be the event space of E.

Let Δ be the class of subsets of S forming the class of all events connected to E. A mapping X of S to R is called a *random variable* or a *stochastic variable* or a *variate*, if for any $x \in R$, the set

$$\{\omega : -\infty < X(\omega) < x,\ \omega \in S\} \in \Delta \tag{5.1.1}$$

i.e., if $\{\omega : -\infty < X(\omega) < x,\ \omega \in S\}$ is an event connected to E. The range of the mapping $X : S \to R$ is called the *spectrum* of the random variable X.

It can be seen that the sets

$\{\omega : X(\omega)=x,\ \omega \in S\}$, $\{\omega : -\infty < X(\omega) < x,\ \omega \in S\}$,

$\{\omega : x < X(\omega) < \infty,\ \omega \in S\}$, $\{\omega : X(\omega) \geqslant c,\ \omega \in S\}$,

$\{\omega : a < X(\omega) < b,\ \omega \in S\}$, etc.

are all subsets (of S) belonging to Δ and so are events connected to E.

Answers

1. (*i*) $\binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2=\frac{125}{3888}$;

(*ii*) $\binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2+\binom{5}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)+\binom{5}{5}\left(\frac{1}{6}\right)^5=\frac{23}{648}$.

2. (*i*) $\binom{8}{3}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3=\frac{7}{32}$; (*ii*) $\binom{8}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3=\frac{7}{32}$;

(*iii*) $1-\{{}^8C_0\left(\frac{1}{2}\right)^8+{}^8C_1\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^7\}=1-\frac{9}{2^8}=\frac{247}{256}$.

3. If A_i denotes the event 'A wins i games' the required probability$=P(A_6)+P(A_7)+P(A_8)=\sum_{r=6}^{8}\binom{8}{r}\left(\frac{2}{3}\right)^r\left(\frac{1}{3}\right)^{8-r}$.

4. The probability of any member of the population being colour-blind $=\frac{5}{100}=\frac{1}{20}$.

$\therefore$ the required probability$=1-\binom{100}{0}\left(\frac{19}{20}\right)^{100}-\binom{100}{1}\left(\frac{1}{20}\right)\left(\frac{19}{20}\right)^{99}$.

5. $\binom{150}{3}\left(\frac{1}{100}\right)^3\left(\frac{99}{100}\right)^{147}$ 7. $\sum_{r=2}^{10}\binom{10}{r}(\cdot 2)^r(\cdot 8)^{10-r}$

10. 6. 11. ${}^4C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=\frac{3}{8}$.

18. $\frac{10!}{4!\,2!\,(1!)^4}\left(\frac{1}{6}\right)^{10}$ by multinomial law.

21. $\sum_{i=a}^{a+b-1}\binom{a+b-1}{i}p^iq^{a+b-1-i}$. 25. 3.

CHAPTER V

PROBABILITY DISTRIBUTIONS

5·1. Random Variables :

We know that the results of a random experiment are objects which are not necessarily real numbers and so it will not be easy to develop a mathematical theory dealing with these objects. Now it is possible to establish a correspondence between outcomes of a random experiment and a given set of real numbers by means of a mapping defined on the event space of the given random experiment as domain and by such correspondence it will be convenient to use algebra and analysis of real numbers in the development of the mathematical theory of probability. With this motivation we introduce the concept of a random variable connected to the event space of any given random experiment.

Definition of a random variable :

Let E be a random experiment and S be the event space of E.

Let $\mathcal{A}$ be the class of subsets of S forming the class of all events connected to E. A mapping X of S to R is called a *random variable* or a *stochastic variable* or a *variate*, if for any $x\in R$, the set

$$\{\omega : -\infty < X(\omega) < x,\ \omega\in S\}\in\mathcal{A} \tag{5.1.1}$$

i.e., if $\{\omega : -\infty < X(\omega) < x,\ \omega\in S\}$ is an event connected to E. The range of the mapping $X : S\to R$ is called the *spectrum* of the random variable X.

It can be seen that the sets

$\{\omega : X(\omega)=x,\ \omega\in S\}$, $\{\omega : -\infty < X(\omega) < x,\ \omega\in S\}$,

$\{\omega : x < X(\omega) < \infty,\ \omega\in S\}$, $\{\omega : X(\omega) > c,\ \omega\in S\}$,

$\{\omega : a < X(\omega) < b,\ \omega\in S\}$, etc.

are all subsets (of S) belonging to $\mathcal{A}$ and so are events connected to E.

In future, we will write the events

$\{\omega : -\infty < X(\omega) \leqslant x, \omega \in S\}$, $\{\omega : X(\omega)=x, \omega \in S\}$,

$\{\omega : a < X(\omega) \leqslant b, \omega \in S\}$ and so on, in short as

$(-\infty < X \leqslant x)$, $(X=x)$, $(a < X \leqslant b)$ respectively and so on.

Ex. A coin is tossed twice. Here,

$$S=\{\omega_1=(H, H), \omega_2=(H, T), \omega_3=(T, H), \omega_4=(T, T)\}.$$

A mapping $X: S \to R$ is defined as follows :

$X(\omega_i)=k$, where k is the number of heads, $i=1, 2, 3, 4$.

Then $X(\omega_1)=2$, $X(\omega_2)=X(\omega_3)=1$, $X(\omega_4)=0$. Here X is a random variable defined in the domain S and the spectrum (range) of X is $\{0, 1, 2\}$. Here, according to our notation $(X=0)$ represents the event $\{(T, T)\}$, $(0 \leqslant X \leqslant 2)$ is a certain event and $(1 < X < 2)$ represents the impossible event O.

The above random variable $X: S \to R$ is also described in the following manner. The random variable X, in this case, defined on S denotes the total number of heads in two tosses of the coin. Later, we shall often use this convention of description of a random variable.

5·2. Distribution Function.

Let $P: \Delta \to R$ be a probability function, where Δ is the class of subsets (of S) forming the class of events. We remember that, the ordered 3 tuple (S, Δ, P) is called a *probability space.*

Let X be a random variable defined on the event space S connected to a random experiment E. The *distribution function* of the random variable X with respect to the probability space (S, Δ, P) is a real valued function $F(x)$ of a real variable x, defined in $(-\infty, \infty)$, where

$$F(x)=P(-\infty < X \leqslant x), \text{ for all } x \in (-\infty, \infty). \tag{5.2.1}$$

It is evident that the range of the distribution function is a subset of $[0, 1]$.

Properties of Distribution Function :

I. $0 \leq F(x) \leq 1$, *for all* $x \in (-\infty, \infty)$. (5.2.2)

Since $0 \leq P(-\infty < X \leq x) \leq 1$ for all $x \in (-\infty, \infty)$, we have $0 \leq F(x) \leq 1$ for all $x \in (-\infty, \infty)$.

II. $P(a < X \leq b)=F(b)-F(a)$. (5.2.3)

The events $(-\infty < X \leq a)$ and $(a < X \leq b)$ are mutually exclusive and

$$(-\infty < X \leqslant a)+(a < X \leqslant b)=(-\infty < X \leqslant b).$$

$\therefore$ $P(-\infty < X \leqslant a)+P(a < X \leqslant b)=P(-\infty < X \leqslant b)$

or, $F(a)+P(a < X \leqslant b)=F(b)$

or, $P(a < X \leqslant b)=F(b)-F(a)$.

III. *F(x) is a monotonically increasing function.* (5.2.4)

Let $x_2 > x_1$. Then by (5.2.3),

$$P(x_1 < X \leqslant x_2)=F(x_2)-F(x_1).$$

But $P(x_1 < X \leqslant x_2) \geqslant 0$ and so $F(x_2) \geqslant F(x_1)$ whenever $x_2 > x_1$.

Hence $F(x)$ is a monotonically increasing function.

IV. *F(x) is continuous to the right at every point a, i.e.,*

$$\lim_{x\to a+0} F(x)=F(a), \text{ or, } F(a+0)=F(a). \tag{5.2.5}$$

Let us construct a sequence of events $\{A_n\}$, defined by

$A_n=\left\{\omega : a < X(\omega) \leqslant a+\frac{1}{n}, \omega \in S\right\}$, where n is a positive integer and S is the corresponding event space. We see that $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots\ldots$ *i.e.,* $\{A_n\}$ is a monotonically decreasing sequence of events.

$$\therefore \text{ by (3.6.1), } P\left(\lim_{n\to\infty} A_n\right)=\lim_{n\to\infty} P(A_n). \tag{5.2.6}$$

Now $\lim_{n\to\infty} A_n=\prod_{n\to\infty}^{\infty} A_n$. If the intersection is not empty, then there exists at least one point ω_1 (say) such that $\omega_1 \in A_n$ for all $n \in N$.

$$a < X(\omega_1) \leqslant a+\frac{1}{n} \text{ for all } n \in N$$

i.e., $\quad a < c \leqslant a+\frac{1}{n}$ for all n, where $X(\omega_1)=c$.

Now $c-a > 0$. Then consider the numbers 1 and $c-a (> 0)$. By the Archimedean property, there will exist a positive integer k such that $k(c-a) > 1$, *i.e.,* $c > a+\frac{1}{k}$. But $c \leqslant a+\frac{1}{n}$ for all $n \in N$. Hence we get a contradiction.

This implies that $\prod\limits_{n\to\infty} A_n = O$, the impossible event.

Then from (5.2.6),

$$P(O) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P\left(a < X \leqslant a+\frac{1}{n}\right)$$

$$= \lim_{n\to\infty} \left\{F\left(a+\frac{1}{n}\right) - F(a)\right\}$$

$$= \lim_{n\to\infty} F\left(a+\frac{1}{n}\right) - F(a).$$

Now $F(x)$ being monotone increasing, $\lim\limits_{x\to a+0} F(x)$ exists finitely and

$$\lim_{n\to\infty} F\left(a+\frac{1}{n}\right) = \lim_{x\to a+0} F(x) = F(a+0).$$

$\therefore \quad o = F(a+o) - F(a)$

i.e., $F(a+o) = F(a)$.

Hence, $F(x)$ is continuous to the right at every point a.

V. *For any real constant* a,

$$F(a) - \underset{x\to a-0}{Lt} F(x) = F(a) - F(a-0) = P(X=a) \qquad (5.2.7)$$

We consider the following sequence of events $\{A_n\}$ defined by

$$A_n = \left\{\omega : a-\frac{1}{n} < X(\omega) \leqslant a,\ \omega \in S\right\},\ n \text{ is a positive integer}$$

and S is the corresponding event space.

Evidently, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, i.e., the sequence $\{A_n\}$ is monotonically decreasing and so

$$\underset{n\to\infty}{Lt} A_n = \prod_{n=1}^{\infty} A_n.$$

We now show that $\prod\limits_{n=1}^{\infty} A_n = \{\omega : X(\omega) = a,\ \omega \in S\}$.

We note that for any $\omega \in S$, where $X(\omega) = a$, $\omega \in A_n$, for all n.

$\therefore \quad \omega \in \prod\limits_{n=1}^{\infty} A_n$, where $X(\omega) = a$.

If possible, let $\omega_1 \in \prod\limits_{n=1}^{\infty} A_n$, where $X(\omega_1) \neq a$.

Then $\omega_1 \in A_n$ for all n.

$\therefore \quad a - \frac{1}{n} < X(\omega_1) < a$ for all n

or, $a - \frac{1}{n} < c < a$ for all n, where $X(\omega_1) = c$.

Here $a - c > 0$. Then considering the two numbers $a-c$ and 1, by Archimedean property there will exist a positive integer p such that $p\cdot(a-c) > 1$, i.e., $c < a - \frac{1}{p}$ and this is a contradiction to the fact that $a - \frac{1}{n} < c$ for all n.

Hence, $\prod\limits_{n=1}^{\infty} A_n = \{\omega : X(\omega) = a,\ \omega \in S\} = (X = a)$.

Now since $\{A_n\}$ is monotonically decreasing, by (3.6.1),

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

or, $$P(\prod_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$$

or, $$P(X=a) = \lim_{n\to\infty} P\left(a - \frac{1}{n} < X \leqslant a\right)$$

$$= \lim_{n\to\infty} \left\{F(a) - F\left(a - \frac{1}{n}\right)\right\}$$

$$= F(a) - \lim_{n\to\infty} F\left(a - \frac{1}{n}\right).$$

Now $F(x)$ is monotonically increasing.

$\therefore \quad \lim\limits_{x\to a-0} F(x)$ exists finitely and

$$\lim_{n\to\infty} F\left(a - \frac{1}{n}\right) = \lim_{x\to a-0} F(x).$$

$$\therefore \quad P(X=a) = F(a) - \lim_{x\to a-0} F(x)$$

or, $F(a) - F(a-0) = P(X=a)$.

VI. $F(\infty) = 1$, where $F(\infty) = \lim\limits_{x\to-\infty} F(x)$ (5.2.8)

We consider the sequence of events $\{A_n\}$ defined by

$$A_n = \{\omega : -\infty < X(\omega) \leqslant n\ ;\ \omega \in S\}$$

where n is a positive integer and S is the corresponding event space.

Then $A_1 \subseteq A_2 \subseteq A_3 \subseteq A_4 \subseteq \ldots\ldots$

i.e., $\{A_n\}$ is a monotonically increasing sequence of events.

$$\therefore \quad \lim_{n\to\infty} A_n = \sum_{n=1}^{\infty} A_n.$$

We now show that $\sum_{n=1}^{\infty} A_n = S$.

$$\text{Since} \quad A_n \subseteq S, \text{ for all } n, \sum_{n=1}^{\infty} A_n \subseteq S. \tag{5.2.9}$$

Now let ω be any element of S and $X(\omega)=d$, a real number. Considering the real numbers d and $1(>0)$, by Archimedean property, there exists a positive integer m such that $m \,.\, 1 > d$ *i.e.,* $-\infty < d < m$, or, $-\infty < X(\omega) < m$ and this implies that $\omega \in A_m$, *i.e.,* $\omega \in \sum_{n=1}^{\infty} A_n$.

Thus $\omega \in S$ implies that $\omega \in \sum_{n=1}^{\infty} A_n$.

$$\therefore \quad S \subseteq \sum_{n=1}^{\infty} A_n. \tag{5.2.10}$$

From (5.2.9) and (5.2.10), $\sum_{n=1}^{\infty} A_n = S$.

$$\therefore \quad P\left(\sum_{n=1}^{\infty} A_n\right) = P(S) = 1. \tag{5.2.11}$$

Now $P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$, by (3.7.1)

or, $P\left(\sum_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(-\infty < X \leqslant n)$.

or, $\lim_{n\to\infty} F(n) = 1$, by (5.2.11).

Now $F(x)$ being monotonically increasing and bounded, $\underset{x\to\infty}{Lt} F(x)$ exists finitely and

$$\underset{x\to\infty}{Lt} F(x) = \underset{n\to\infty}{Lt} F(n) = 1.$$

Hence it is proved that $F(\infty)=1$.

$$\text{VII.} \quad F(-\infty)=0, \text{ where } F(-\infty) = \underset{x\to-\infty}{Lt} F(x) \tag{5.2.12}$$

We consider the sequence of events $\{A_n\}$ defined by

$A_n = \{\omega : -\infty < X(\omega) \leqslant -n, \omega \in S\}$, n is a positive integer and S is the corresponding event space.

Here $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots\ldots$

i.e., $\{A_n\}$ is a monotonically decreasing sequence of events.

$$\therefore \quad \underset{n\to\infty}{Lt} A_n = \prod_{n=1}^{\infty} A_n.$$

We show that $\prod_{n=1}^{\infty} A_n$ is an empty set, *i.e.,* it is the impossible event O.

If possible let $\omega \in A_n$ for all n.

$\therefore \quad -\infty < X(\omega) \leqslant -n$ for all n,

i.e., $-\infty < c \leqslant -n$ for all n, where $X(\omega)=c$.

This implies that $n \leqslant -c$ for all n, *i.e.,* the set of positive integers is bounded above, which is a contradiction.

$$\therefore \quad \prod_{n=1}^{\infty} A_n = O.$$

Now, by (3.6.1), $\underset{n\to\infty}{Lt} P(A_n) = P(\underset{n\to\infty}{Lt} A_n)$

or, $\underset{n\to\infty}{Lt} P(-\infty < X \leqslant -n) = P(O) = 0$

or, $\underset{n\to\infty}{Lt} F(-n) = 0$.

Now $F(x)$ is a monotonically increasing and bounded function. Hence $\underset{x\to-\infty}{Lt} F(x)$ exists finitely.

$$\underset{x\to-\infty}{Lt} F(x) = \underset{n\to\infty}{Lt} F(-n) = 0,$$

i.e., $F(-\infty)=0$.

VIII. *The set of points of discontinuity of a distribution function is at most enumerable.*

We know that every monotonic function can have at most a countable set of points of discontinuity. Since every distribution function is monotonic, the property follows.

Remark: (*a*) From properties I-VII we conclude that the distribution function $F(x)$ is a monotonic non-decreasing bounded function such that

(*i*) $F(-\infty)=0$

(*ii*) $F(\infty)=1$

(*iii*) it is continuous to the right at all points

(*iv*) it is discontinuous to the left at every point $x=a$, if $P(X=a)>0$ and the discontinuity being a jump discontinuty, the height of the jump (or saltus) is equal to $P(X=a)$.

(*b*) The converse of the remark (*a*) is also true and so we conclude the following (without proof):

Any function $F(x)$ with domain $(-\infty, \infty)$ and range a subset of $[0, 1]$ is a distribution function of a random variable with respect to a probability space (S, Δ, P) if and only if $F(x)$ is such that (*i*) $F(-\infty)=0$, (*ii*) $F(\infty)=1$, (*iii*) $F(x)$ is monotonically non-decreasing and bounded, (*iv*) $F(x)$ is continuous to the right at all points, (*v*) $F(x)$ is discontinuous to the left at every point $x=a$, if $P(X=a)>0$.

(*c*) The curve $y=F(x)$ is called the *distribution curve* of the corresponding random variable X. It is evident that the distribution curve lies between $y=0$, $y=1$.

Probability distribution and the concept of probability mass.

If the distribution function $F(x)$ of a random variable X be known, then for any a, b $(a<b)$, the probability of the event $(a<X\leqslant b)$ can be determined. So the distribution function $F(x)$ gives the distribution of probabilites of various events and so we say that $F(x)$ determines the *probability distribution* of the random variable X. Then the problem of determination of the probability distribution of X is the same as the problem of finding the distribution function $F(x)$ of X.

From the properties of the distribution function proved in § 5.2 (I-VII), it will be possible to make an analogy with 'probability of an event' and 'mass of a particle or of a system of particles.' The aforesaid analogy can be done as follows: We assume that a certain amount of matter is distributed on a given straight line (on which x is measured) in such a way that the total mass of the matter distributed from $-\infty$ up to the point $x=a$ is equal to $F(a)$, where $F(x)$ is the distribution function of the random variable X. Then the property '$F(\infty)=1$' implies immediately that the total mass of matter distributed on the line is 1 unit. The property '$P(a<X\leqslant b)=F(b)-F(a)$' reflects that the probability of the event $(a<X\leqslant b)$ is equal to the mass of the matter distributed on the semi-closed interval $(a, b]$. The relation '$P(X=a)=F(a)-F(a-0)$' shows that the probability of the event $(X=a)$ can be interpreted as the mass of a particle placed at the point $x=a$.

The hypothetical distribution of mass described above is called the *probability mass* and in many situations it will be convenient to think probability in terms of mass by the aforesaid analogy where the probabilty of an event is identified with the mass of a certain amount of matter.

We shall restrict our discussion to two types (unless otherwise stated) of random variables, namely discrete and continuous which will be explained in the following sections.

5.3. Discrete Distribution. Probability Mass Function (p.m.f.)

A random variable X defined on an event space S is said to be discrete if the spectrum of X is at most countable, *i.e.*, if the spectrum is finite or countably infinite. In this case the probability distribution of X will be called a *discrete distribution.*

Let the spectrum of X be $\{x_i : i=0, \pm 1, \pm 2, \ldots\ldots\}$, where $\ldots\ldots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \ldots\ldots$

Let $P(X=x_i)=f_i$, x_i being a spectrum point. A function $f: R \to [0, 1]$ is defined as follows:

$$f(x)=f_i, \text{ if } x=x_i, \text{ which is a point of the spectrum,} \\ =0, \text{ elsewhere.} \tag{5.3.1}$$

The function f defined above is called the *probability mass function* (p. m. f.) of the random variable X.

The *distribution function* $F(x)$ *of a discrete random variable* X is given by:

$$F(x)=\sum_{x_j \leq x_i}^{i} P(X=x_j)=\sum_{j=-\infty}^{i} f_j, \text{ if } x_i \leq x < x_{i+1}$$
$$(i=0, \pm 1, \pm 2, \ldots\ldots). \quad (5.3.2)$$

Thus $F(x)$ is a step function which remains constant over every interval in between two consecutive spectrum points, has a jump discontinuity at each spectrum point x_i, the height of the jump at each point being $f_i=P(X=x_i)$. It is continuous to the right but discontinuous to the left at each spectrum point.

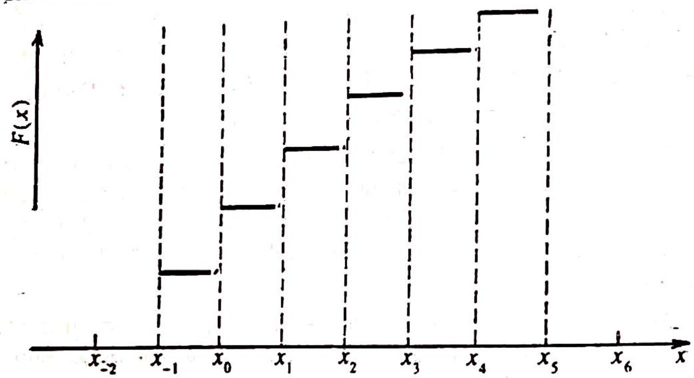


Fig. 5.3.1 Distribution Function of a Discrete Distribution.

5.4. Some important results on discrete distributions.

I. $\sum_{j=-\infty}^{\infty} f_j=1.$ (5.4.1)

By (5.2.8.), $F(\infty)=1$.

But from (5.3.2), $F(\infty)=\sum_{j=-\infty}^{\infty} f_j$

$\therefore \sum_{j=-\infty}^{\infty} f_j=1.$

II. *At each non-spectrum point 'a',*
$P(X=a)=0.$ (5.4.2)

We have $P(X=a)=F(a)-F(a-0)$, by (5.2.7). Now a being a non-spectrum point, there exist step points x_k and x_{k+1} such that $x_k < a < x_{k+1}$.

Now $F(x)=\sum_{j=-\infty}^{k} f_j$, when $x_k \leq x < x_{k+1}$.

Since $x_k < a < x_{k+1}$, $F(a)=\sum_{j=-\infty}^{k} f_j$.

Also $F(a-0)= \underset{x \to a-0}{Lt} F(x)$, where $x_k < x < a < x_{k+1}$,

$$=\sum_{j=-\infty}^{k} f_j.$$

$\therefore$ $F(a)=F(a-0)$ and hence $P(X=a)=0$, whenever a is a non-spectrum point.

III. $P(a < X \leq b)=\sum_{a < x_i \leq b} f_i$. (5.4.3)

Given any half open interval $a < x \leq b$, we find all spectrum points x_i such that $a < x_i \leq b$.

Now $P(a < X \leq b)=F(b)-F(a)$, by (5.2.3)

$$=\sum_{a < x_i \leq b} f_i$$

where the summation is taken over all values of i such that $a < x_i \leq b$.

IV. *Let X be a discrete random variable with x_i $(i=0, \pm 1, \pm 2, \ldots)$ as spectrum points. If $P(X=x_i)=f_i$ be given, for $i=0, \pm 1, \pm 2, \ldots$, the distribution function can be determined and conversely.*

Let $\{x_i : i=0, \pm 1, \pm 2, \ldots\ldots\}$ be the spectrum of X, where $\ldots\ldots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \ldots\ldots$, and let $P(X=x_i)=f_i$.

We define a function F as follows :

$$F(x)=\sum_{j=-\infty}^{i} f_j, \text{ for } x_i \le x < x_{i+1}, \ (i=0, \pm 1, \pm 2, \ldots).$$

Then $F(x)$ is the distribution function of X.

Conversely, let the distribution function $F(x)$ of a discrete random variable X be given, where the spectrum of X is the set $\{x_i : i=0, \pm 1, \pm 2, \ldots\}$.

Then $P(X=x_i)$ denoted as f_i is given by $f_i=P(X=x_i)=F(x_i)-F(x_i-0)$ for $i=0, \pm 1, \pm 2, \ldots$

We now discuss some important discrete distributions.

5.5. Important discrete distributions.

I. *Binomial (n, p) Distribution.*

A discrete random variable X having the set $\{0, 1, 2, \ldots, n\}$ as the spectrum, is said to have *binomial distribution* with parameters n, p if the p. m. f. of X is given by,

$$f(x)=\binom{n}{x}p^x(1-p)^{n-x}, \text{ for } x=0, 1, 2, \ldots, n$$
$$=0, \text{ elsewhere,}$$

where n is a positive integer and $0 < p < 1$.

We now give one example of binomial distribution from real life situation. Let E_n be the resulting compound experiment arising from n (a positive integer) Bernoulli trials, where $p(0 < p < 1)$ is the probability of success in each trial. If we are interested only in the number of successes, then the event space corresponding to E_n is the finite set $\{0, 1, 2, \ldots, n\}=S$ (say).

A random variable X is defined on S as follows :

$$X(i)=i \quad \text{where} \quad i \in S.$$

Then X is a discrete random variable where the probability mass function $f(x)=P(X=x)$ is given by

$$f(x)=\binom{n}{x}p^x(1-p)^{n-x} \text{ for } x=0, 1, 2, \ldots n$$
$$=0, \text{ elsewhere.} \tag{5.5.1}$$

(5.5.1) shows that X has binomial (n, p) distribution.

II. *Poisson μ Distribution.*

A discrete random variable X having the enumerable set $\{0, 1, 2, \ldots\}$ as the spectrum, is said to have *Poisson distribution* with parameter $\mu(>0)$, if the p.m.f. is given by

$$f(x)=\frac{e^{-\mu}\mu^x}{x!}, \text{ for } x=0, 1, 2, \ldots$$
$$=0, \text{ elsewhere.}$$

Let us now give an example of Poisson distribution from real life problems. If X be the random variable denoting the number of telephone calls in a given interval $(0, t)$, satisfying the conditions in a Poisson process (see § 5.11), then X is a discrete variate whose spectrum is the enumerable set $\{0, 1, 2, \ldots\}$, the corresponding probability mass function $f(x)=P(X=x)$ is given by

$$f(x)=\frac{e^{-\lambda t}(\lambda t)^x}{x!} \text{ for } x=0, 1, 2, \ldots$$
$$=0, \text{ elsewhere.} \tag{5.5.2}$$

(λ is the average number of calls per unit time).

(5.5.2) shows that X has Poisson distribution with parameter λt.

III. *Geometric Distribution.*

A discrete random variable X is said to have *geometric distribution* with parameter $p(0 < p < 1)$, if the probability mass function $f(x)=P(X=x)$ is given by

$$f(x)=pq^x \text{ for } x=0, 1, 2, 3, \ldots$$
$$=0, \text{ elsewhere,} \tag{5.5.3}$$

where $q=1-p$.

We observe that $\displaystyle\sum_{x=0}^{\infty} pq^x=1$.

In an infinite sequence of Bernoulli trials, if X denotes the number of failures preceding the first success, then X has the above distribution.

IV. *Negative Binomial Distribution.*

In an infinite sequence of Bernoulli trials let X be the random variable denoting the number of failures that precede the rth

success, where $r \geq 1$ is a fixed positive integer. Then X is a discrete random variable, whose probability mass function $f(x)=P(X=x)$ is given by

$$f(x)=\binom{x+r-1}{x}(1-p)^x p^r \text{ for } x=0, 1, 2, \ldots\ldots \tag{5.5.4}$$
$$=0, \text{ elsewhere.}$$

This random variable X is said to have *negative binomial distribution* with parameters r, p.

It can be verified that $\sum_{x=0}^{\infty} f(x)=1.$

V. *Hypergeometric Distribution.*

Let N_1, N_2 be two given positive integers and let $N_1+N_2=N$.

A random variable X is said to have *hypergeometric disribution,* with parameters N_1, N_2 and n, if the probability mass function $f(x)=P(X=x)$ is given by

$$f(x)=\frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \tag{5.5.5}$$

where max. $(0, n-N_2) \leqslant x \leqslant$ min. (N_1, n), x is a non-negative integer and n is a positive integer $(n \leqslant N)$.

Let a box contain N_1 white and N_2 black balls. n balls are drawn at random from the box without replacement. If X be the random variable denoting the number of white balls drawn, then X has the above distribution.

We note that

$$\sum_{k=0}^{n}\binom{N_1}{k}\binom{N_2}{n-k}=\binom{N_1+N_2}{n}.$$

$$\therefore \sum_{x} P(X=x)=\frac{1}{\binom{N}{n}}\sum_{x}\binom{N_1}{x}\binom{N_2}{n-x}=\frac{\binom{N}{n}}{\binom{N}{n}}=1.$$

5.6. Continuous Random Variable.

Let X be a random variable defined on an event space S. Let $F(x)$ be the distribution function of X. Then the random variable X is said to be continuous if

(*i*) the distribution function $F(x)$ is continuous for all real values of x,

(*ii*) and for any two real numbers a, b $(a < b)$, $\frac{d}{dx}F(x)=F'(x)$ is continuous in $[a, b]$ except for at most a finite number of discontinuities (which may include points of infinite discontinuity) and $\int_a^b F'(x)\,dx$ is convergent.

Alternative definition of continuous variate.

A random variable X defined on the event space S is said to be a continuous random variable if there exists a non-negative real valued function $f(x)$ such that (*i*) $f(x)$ is integrable in $(-\infty, \infty)$ and (*ii*) the distribution function $F(x)$ of X is given by $F(x)=\int_{-\infty}^{x} f(t)\,dt$ for any real x.

The equivalence of the two definitions will follow from (5.8.3) and note (*d*) of § 5·8. If X is a *continuous random variable,* then the probability distribution of X is called a *continuous distribution.*

5.7. Probability Density Fuuction (p.d.f) of a Continuous Distribution.

In case of a continuous distribution, we denote $F'(x)$ by $f(x)$, where $f(x)$ is called the *probability density function* (p.d.f.) of X, $F(x)$ being the distribution function of X. From definition, the density function is continuous in any finite interval $[a, b]$ except for at most finite number of points of discontinuity. We note that $F'(x)$ may not be defined values of x and consequently $f(x)$ may be undefined at some points.

In the alternative definition of a continuous variate X, the non-negative real valued function $f(x)$ is called a probability density function of X. Here from the relation $F(x)=\int_{-\infty}^{x} f(t)dt$ we get $F'(x)=f(x)$ at a point of continuity x of $f(x)$.

5.8. Some Important Results on the probability density function f and the corresponding distribution function F of a continuous variate X.

I. *$f(x) \geq 0$ for all x where $f(x)$ is defined.* (5.8.1)

We know that $F(x)$ is a monotonic increasing function. So $F'(x) \geq 0$ whenever $F'(x)$ exists.

$\therefore$ $f(x) \geq 0$ for all x, where $f(x)$ is defined.

II. $P(a < X \leq b) = \int_a^b f(x)\, dx.$ (5.8.2)

We have $P(a < X \leq b) = F(b) - F(a)$.

Now $F'(x) = f(x)$ is continuous in $[a, b]$ except for at most a finite numbe of discontinuities and So, we have $\int_a^b F'(x)\, dx$ is convergent.

Then, $\int_a^b f(x)\, dx = F(b) - F(a) = P(a < X \leq b)$.

III. $F(x) = \int_{-\infty}^{x} f(t)\, dt.$ (5.8.3)

We have, by (5.8.2),

$$P(a < X \leq x) = \int_a^x f(t)\, dt$$

or, $F(x) - F(a) = \int_a^x f(t)\, dt.$

Proceeding to the limit $a \to -\infty$, we get

$$F(x) - \underset{a \to -\infty}{Lt} F(a) = \underset{a \to -\infty}{Lt} \int_a^x f(t)\, dt$$

or, $F(x) - F(-\infty) = \int_{-\infty}^{x} f(t)\, dt$

or, $F(x) = \int_{-\infty}^{x} f(t)\, dt$, since $F(-\infty) = 0$.

IV. $\int_{-\infty}^{\infty} f(x)\, dx = 1.$ (5.8.4)

We have, by (5.8.3),

$$\int_{-\infty}^{x} f(t)\, dt = F(x).$$

$\therefore$ proceeding to the limit $x \to \infty$,

$$\underset{x \to \infty}{Lt} \int_{-\infty}^{x} f(t)\, dt = \underset{x \to \infty}{Lt} F(x) = F(\infty) = 1$$

or, $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

V. *$P(X = a) = 0$, where a is a given constant.* (5.8.5)

We have $F(a) - F(a-0) = P(X = a)$.

Now X being a continuous random variable, $F(x)$ is continuous for all x. Hence $F(x)$ is continuous at $x = a$.

$\therefore$ $\underset{x \to a-0}{Lt} F(x) = F(a)$,

i.e., $F(a-0) = F(a)$.

$\therefore$ $P(X = a) = F(a) - F(a-0) = 0$,

i.e., $P(X = a) = 0$ for any real constant a.

Note: (*a*) We see that the distribution function of a continuous random variable X is completely determined by the corresponding probability density function $f(x)$, using (5.8.3). So the probability distribution of a continuous random variable is completely determined by the corresponding density function $f(x)$.

(*b*) We observe that the probability density function defined in the two ways mentioned in § 5·7 may differ at some points but they will determine the same distribution function $F(x)$ of a continuous random variable X. Further we observe that if the values of the p.d.f. (in any definition) be altered at finite number of points or if the p.d f. be defined arbitrarily at finite number of points where it is undefined, then the corresponding distribution function $F(x)$ is not altered.

(*c*) We know that $P(O) = 0$. If, however $P(A) = 0$, we cannot conclude that A is an impossible event. In this case, we say that A is *stochastically impossible.* We now give an example to show that 'an event may be stochastically impossible but not impossible.'

Let E be the random experiment of selecting a number at random from the open interval (0, 9). Let X be the random variable denoting the number chosen. Then an event '$X = 6$' is not an impossible event. But it can be shown that X is a continuous

random variable and so $P(X=6)=0$. So the event '$X=6$' is a stochastically impossible event but not an impossible event.

(d) Every non-negative real valued piecewise continuous function $f(x)$ that is integrable in $(-\infty, \infty)$ and for which $\int_{-\infty}^{\infty} f(x)\,dx=1$, is the probability density function of a continuous distribution. (5.8.6)

It is sufficient to show that there exists a distribution function $F(x)$ corresponding to $f(x)$.

We define a function F given by $\int_{-\infty}^{x} f(t)\,dt=F(x)$, (5.8.7)

which exists for any real x.

Now $F(x)=\int_{-\infty}^{x} f(t)\,dt=\underset{B\to-\infty}{Lt}\int_{B}^{x} f(t)\,dt$

$$=\underset{B\to-\infty}{Lt}\left[\int_{-\infty}^{x} f(t)\,dt-\int_{-\infty}^{B} f(t)\,dt\right]$$

$$=\underset{B\to-\infty}{Lt}[F(x)-F(B)]$$

$$=F(x)-F(-\infty)].$$

$\therefore\ F(-\infty)=0.$

Also from (5.8.7) proceeding to the limit $x\to\infty$,

$$F(\infty)=\int_{-\infty}^{\infty} f(t)\,dt=1, \text{ by the given condition.}$$

Finally, let $x_1, x_2 \in R$, where $x_2 > x_1$.

Then $F(x_2)=\int_{-\infty}^{x_2} f(t)\,dt=\int_{-\infty}^{x_1} f(t)\,dt+\int_{x_1}^{x_2} f(t)\,dt$

$$\geq \int_{-\infty}^{x_1} f(t)\,dt=F(x_1),$$

since $f(x)\geq 0$ implies $\int_{x_1}^{x_2} f(t)\,dt \geq 0$, when $x_2 > x_1$.

Thus $F(x_2)\geq F(x_1)$ whenever $x_2 > x_1$. Thus $F(x)$ is monotonically non-decreasing.

Finally, $f(x)$ being piecewise continuous, from (5.8.7) it can be shown that $F(x)$ is continuous everywhere; further at a point of continuity of the function $f(x)$, the derivative of $F(x)$ is equal to $f(x)$, i.e., $F'(x)=f(x)$ and so $F'(x)$ is also piecewise continuous. This proves that $F(x)$ is a distribution function of a continuous distribution. Hence the proposition.

Density curve: The graphical representation of $y=f(x)$ is called the probability *density curve* of the corresponding continuous distribution.

VI. *Probability Differential.*

Let X be a continuous random variable and $\delta x > 0$.

Then $P(x < X \leq x+\delta x)=F(x+\delta x)-F(x)=\delta x F'(\xi)$,

$$\xi=x+\theta\delta x,\ 0<\theta<1,$$

by Lagrange's Mean Value Theorem of Differential Calculus.

$$\therefore\ \underset{\delta x\to 0}{Lt}\frac{P(x < X \leq x+\delta x)}{\delta x}=\underset{\delta x\to 0}{Lt} F'(\xi)=F'(x),$$

if x is a point of continuity of $F'(x)=f(x)$.

$$\therefore\ f(x)=\underset{\delta x\to 0}{Lt}\frac{P(x < X \leq x+\delta x)}{\delta x}$$

$$=\underset{dx\to 0}{Lt}\frac{P(x < X \leq x+dx)}{dx},$$

since $\delta x=dx$, the differential of the variable x.

Henceforth we shall write $f(x)\,dx$ for $P(x < X \leq x+dx)$, which will actually mean $\underset{dx\to 0}{Lt}\frac{P(x < X \leq x+dx)}{dx}=f(x)$.

The expression $P(x < X \leqslant x+dx)$ will always be used in the above limiting sense and so there will be no ambiguity throughout our discussion. The expression $P(x < X \leqslant x+dx)$ which is taken to be equal to $f(x)dx$,

i.e., $f(x)\,dx=P(x < X \leqslant x+dx)$ (5.8.8)

is called the *probability differential* of the continuous random variable X.

We now discuss some important continuous distributions.

5.9. Important Continuous Distributions.

I. *Uniform or Rectangular Distribution.*

A continuous random variable X, is said to follow a uniform

distribution, if its probability density function (p.d.f) is given by

$$f(x)=\frac{1}{b-a},\ a<x<b$$
$$=0,\ \text{elsewhere},\qquad (5.9.1)$$

where a and b are the two parameters of the distribution.

We note that $f(x)\geqslant 0$ for all x.

Also $\int_{-\infty}^{\infty} f(x)\,dx=\int_a^b \frac{dx}{b-a}=1.$

The distribution function $F(x)$ of X is given by

$$F(x)=0,\ -\infty<x<a$$
$$=\frac{x-a}{b-a},\ a\leqslant x\leqslant b$$
$$=1,\ b<x<\infty;\qquad (5.9.2)$$

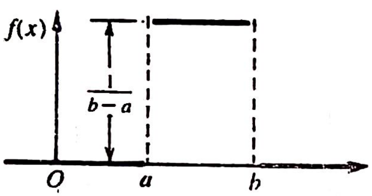


Fig. 5.9.1
Rectangular Density Curve

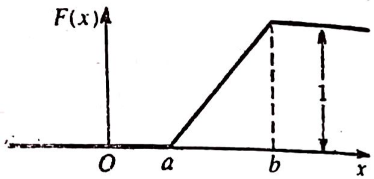


Fig. 5.9.2
Rectangular Distribution Curve

Note: The rectangular distribution gives a useful model for random experiment like 'a point is chosen at random in a given interval'. In this case, we are actually thinking of a random variable X such that the probability of the event 'X lying in any sub-interval' is proportional to the length of the sub-interval, *i.e.*, X is uniformly distributed in the given interval.

II. *Normal* (m, σ) *Distribution.*

A continuous random variable X, having $(-\infty, \infty)$ as the spectrum, is said to follow a *normal distribution* if its probability density function $f(x)$ is given by

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{(x-m)^2}{2\sigma^2}},\quad -\infty<x<\infty \qquad (5.9.3)$$

where $\sigma>0$.

It is a probability distribution with two parameters m, σ and is denoted by $N(m, \sigma)$.

In particular if $m=0$, $\sigma=1$, we say that the corresponding random variable X is a *Standard Normal Variate.*

Since $\frac{1}{\sqrt{2\pi}\,\sigma}>0$ and $e^{-\frac{(x-m)^2}{2\sigma^2}}$ is non-negative, for all values of x, hence $f(x)>0$ for all x

Again $\int_{-\infty}^{\infty} f(x)\,dx=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx$

$$=\frac{1}{\sqrt{2\pi}\,\sigma}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\int_P^Q e^{\frac{-(x-m)^2}{2\sigma^2}}\,dx$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\int_{\frac{P-m}{\sqrt{2}\sigma}}^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\cdot\sqrt{2}\sigma\,dy \text{ where } \frac{x-m}{\sqrt{2}\sigma}=y$$

$$=\frac{1}{\sqrt{\pi}}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\left[\int_{\frac{P-m}{\sqrt{2}\sigma}}^{0} e^{-y^2}\,dy+\int_0^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,dy\right]$$

$$=\frac{1}{\sqrt{\pi}}\ \underset{\substack{Q\to\infty\\ P\to-\infty}}{Lt}\left[-\int_{\frac{-P+m}{\sqrt{2}\sigma}}^{0} e^{-t^2}\,dt+\int_0^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,dy\right]$$

where in the first integral we put $y=-t$

$$=\frac{1}{\sqrt{\pi}}\ \underset{P\to-\infty}{Lt}\int_0^{\frac{-P+m}{\sqrt{2}\sigma}} e^{-t^2}\,dt+\frac{1}{\sqrt{\pi}}\ \underset{Q\to\infty}{Lt}\int_0^{\frac{Q-m}{\sqrt{2}\sigma}} e^{-y^2}\,dy$$

$$=\frac{1}{\sqrt{\pi}}\int_0^{\infty} e^{-t^2}\,dt+\frac{1}{\sqrt{\pi}}\int_0^{\infty} e^{-y^2}\,dy,$$

since the integrals are convergent

$$=\frac{2}{\sqrt{\pi}}\int_0^{\infty} e^{-y^2}\,dy=\frac{2}{\sqrt{\pi}}\left\{\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\right\}=\frac{2}{\sqrt{\pi}}\cdot\frac{\sqrt{\pi}}{2}=1.$$

The distribution function of a normal distribution is given by

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx. \tag{5.9.4}$$

If X is a standard normal variate (*i.e.*, $m=0$, $\sigma=1$), the corresponding density and distribution functions are given by

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \text{ and } \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt. \tag{5.9.5}$$

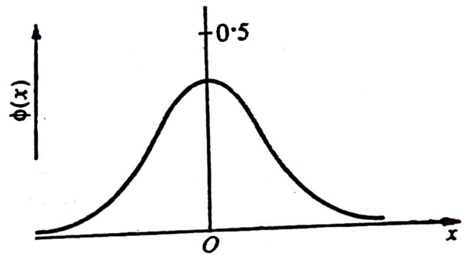


Fig. 5.9.3 Standard Normal Density Curve

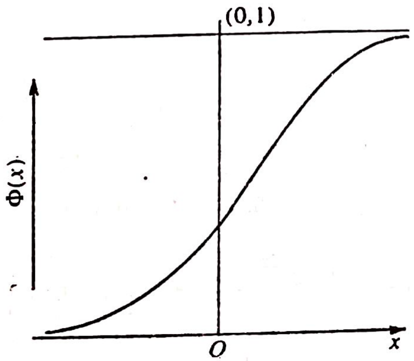


Fig. 5.9.4 Standard Normal Distribution Curve

One of the most important distributions in the theory of probability and statistics is the normal distribution and in the foregoing chapters we study this distribution in detail.

III. *Gamma Distribution.*

A continuous random variable X is said to follw a *Gamma distribution* if its probability density function $f(x)$ is given by

$$f(x)=\frac{e^{-x}\, x^{l-1}}{\Gamma(l)},\ 0<x<\infty,\ l>0$$

$$=0, \text{ elsewhere,} \tag{5.9.6}$$

l being the only parameter of the distribution.

We note that $f(x)\geq 0$ for all x.

Also $\int_{-\infty}^{\infty} f(x)\, dx=\frac{1}{\Gamma(l)}\int_{0}^{\infty} e^{-x}\, x^{l-1}\, dx=\frac{\Gamma(l)}{\Gamma(l)}=1.$

The random variable in question is referred to as $\gamma(l)$ *variate.*

IV. *Beta Distribution of the First Kind.*

A continuous random variable X is said to follow a *Beta distribution* of the first kind, if its probability density function $f(x)$ is given by

$$f(x)=\frac{x^{m-1}(1-x)^{n-1}}{B(m,n)},\ 0<x<1\ ;\ m>0,\ n>0$$

$$=0, \text{ elsewhere.} \tag{5.9.7}$$

The parameters are $m(>0)$, $n(>0)$ and the random variable X is called $\beta_1(m,n)$ *variate.*

Evidently $f(x)\geq 0$ for all x.

Now $\int_{-\infty}^{\infty} f(x)\, dx=\frac{1}{B(m,n)}\int_{0}^{1} x^{m-1}(1-x)^{n-1}\, dx=\frac{B(m,n)}{B(m,n)}=1.$

V. *Beta Distribution of the Second Kind.*

A continuous random variable X is said to follow a *Beta distribution of the second kind* if its probability density function $f(x)$ is given by

$$f(x)=\frac{x^{m-1}}{B(m,n)(1+x)^{m+n}},\ 0<x<\infty\ ;\ m>0,\ n>0$$

$$=0, \text{ elsewhere.} \tag{5.9.8}$$

The random variable X is called $\beta_2(m,n)$ *variate.*

Evidently $f(x)\geq 0$ for all x.

Also since $B(m, n)=\int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}}\,dx$, we get

$$\int_{-\infty}^{\infty} f(x)\,dx=\frac{1}{B(m,n)}\int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}}\,dx=1.$$

VI. *Cauchy Distribution.*

A continuous random variable X is said to follow a *Cauchy distribution* if its probability density function $f(x)$ is given by

$$f(x)=\frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2},\ -\infty < x < \infty \tag{5.9.9}$$

$\lambda > 0$ and μ being the parameters.

It is evident that $f(x) \geqslant 0$ for all x.

Also $\int_{-\infty}^{\infty} f(x)\,dx=1$, can easily be verified.

The distribution function $F(x)$ is given by

$$F(x)=\int_{-\infty}^{x} \frac{1}{\pi}\frac{\lambda}{\lambda^2+(x-\mu)^2}\,dx$$

$$=\frac{\lambda}{\pi}\ \underset{P\to-\infty}{Lt}\int_P^x \frac{dx}{\lambda^2+(x-\mu)^2}$$

$$=\frac{\lambda}{\pi}\ \underset{P\to-\infty}{Lt}\left[\frac{1}{\lambda}\tan^{-1}\frac{x-\mu}{\lambda}\right]_P^x$$

$$=\frac{1}{\pi}\ \underset{P\to-\infty}{Lt}\left(\tan^{-1}\frac{x-\mu}{\lambda}-\tan^{-1}\frac{P-\mu}{\lambda}\right)$$

$$=\frac{1}{\pi}\tan^{-1}\frac{x-\mu}{\lambda}+\frac{1}{2}. \tag{5.9.10}$$

VII. *Exponential Distribution.*

A continuous random variable X is said to follow *exponential distribution if the* probability density function $f(x)$ is given by

$$f(x)=\frac{1}{\lambda}e^{-\frac{x}{\lambda}},\ x>0,\ \lambda>0$$
$$=0, \text{ elsewhere}, \tag{5.9.11}$$

$\lambda > 0$ is the only parameter of the distribution.

We observe that $f(x) \geqslant 0$ for all x and $\int_{-\infty}^{\infty} f(x)\,dx=1$.

5.10. Distinction between Discrete and Continuous Random Variables.

(*a*) A random variable X is discrete when its spectrum is at most an enumerable set, *i.e.*, either finite or countably infinite, whereas in the case of a continuous random variable, the spectrum is usually an interval or union of some intervals.

(*b*) The distribution function of a discrete random variable is a step function, whereas in the case of a continuous random variable, the distribution function $F(x)$ is continuous for all x and in any bounded interval $[a, b]$ $F'(x)$ is continuous except for at most a finite number of points of discontinuity.

(*c*) In the case of a discrete random variable $P(X=a)-0$ if a is not a spectrum point, while in the case of a continuous random variable $P(X=a)=0$ for any real number a.

(*d*) The random variable denoting the number of telephone calls in a given trunk line in a given interval of time is an example of a discrete random variable (see Poisson Process). The random variable denoting the number chosen at random from a given interval, say (4, 7), is an example of a continuous random variable.

5.11. Poisson Process.

Before describing Poisson process, we shall define stochastic process. Let $X(t)$ be a random variable depending on a real variable t, which is usually time in real life problems. A given set of random variables $\{X(t) : t \in D\}$ where D is a given subset of R is called a *stochastic process.*

The random variable denoting the number of persons in a queue at time t for different values of t form a stochastic process. Another example of stochastic process is the set of random variables giving the number of persons injected with a given disease (in a given time t) for different values of t.

THEOREM 5.11.1. *Number of changes of a stochastic process in a given interval of time follow Poisson law under certain conditions.*

Proof: Let $X(t)$ be the random variable denoting the number of changes of a stochastic process in a given interval of time $(0, t)$,

where $t > 0$. We make the following assumptions regarding the probability distribution of $X(t)$:

(*i*) If h be a positive number, then the number of changes in the interval $(0, t)$ and the number of changes in $(t, t+h)$ are independent.

(*ii*) The probability of exactly one change in the interval $(t, t+h)$ is $\lambda h+0(h)$, where $0(h)$ denotes any function of h such that $\underset{h\to 0+}{Lt}\frac{0(h)}{h}=0$ and λ is a positive constant.

(*iii*) The probability of more than one change in $(t, t+h)$ is $0(h)$.

We shall now show that under the above conditions, the distribution of $X(t)$ will follow Poisson law.

Spectrum of $X(t)$ is given by

$$X(t)=i, \quad i=0, 1, 2,\ldots$$

Let E be the random experiment of counting the number of changes in $(0, t)$ and E' be random experiment of counting the number of changes in $(t, t+h)$. By the assumption (1) we can say that E and E' are independent.

Let $P_i(t)$ be the probability of the event

$$`X(t)=i', \quad i=0, 1, 2,\ldots$$

Then $P_i(t+h)$ is the probability of the event

$$`X(t+h)=i'.$$

Case I. $i > 1$.

$P_i(t+h)$=Probability of the event 'i changes in $(0, t+h)$'

$$=P(A_1)+P(A_2)+P(A_3)$$

where A_1 denotes the event '$i-1$ changes in $(0, t)$ and one change in $(t, t+h)$,'

A_2 denotes the event 'i changes in $(0, t)$ and no change in $(t, t+h)$' and

A_3 denotes the event 'more than one change in $(t, t+h)$' such that the total number of changes in $(0, t+h)$ is i.

We note that A_1, A_2, A_3 are mutually exclusive. Now, since the random experiments E and E' are independent, we have

$$P(A_1)=P_{i-1}(t)\times\{\lambda h+0(h)\}$$
$$P(A_2)=P_i(t)\times\{1-\lambda h-0(h)-0(h)\}$$

and $P(A_3)$ can be expressed as $0(h)$, by assumption (*iii*).

$$\therefore\quad P_i(t+h)=P_{i-1}(t)\times\{\lambda h+0(h)\}+P_i(t)-\lambda hP_i(t)$$
$$-P_i(t)[0(h)+0(h)]+0(h),$$

[Here, $0(h)+0(h)$, $P_i(t)\times 0(h)$, $-P_i(t)[0(h)+0(h)]$ can all be replaced by $0(h)$]

$$=\lambda hP_{i-1}(t)+0(h)+P_i(t)-\lambda hP_i(t)+0(h)+0(h)$$
$$=\lambda hP_{i-1}(t)+P_i(t)-\lambda hP_i(t)+0(h).$$

$$\therefore\quad \frac{P_i(t+h)-P_i(t)}{h}=\lambda P_{i-1}(t)-\lambda P_i(t)+\frac{0(h)}{h}$$

Hence proceeding to the limit $h\to 0+$, we get

$$P_i'(t)=\lambda\{P_{i-1}(t)-P_i(t)\} \text{ for } i > 1. \tag{5.11.1}$$

Case II. $i=1$.

In this case $P_1(t+h)=P(B_1)+P(B_2)$, where B_1 denotes the event 'exactly one change in $(0, t)$ and no change in $(t, t+h)$', and B_2 denotes the event 'no change in $(0, t)$ and exactly one change in $(t, t+h)$'.

So, E and E' being independent,

$$P_1(t+h)=P_1(t)\{1-\lambda h-0(h)-0(h)\}+P_0(t)\{\lambda h+0(h)\}$$
$$=P_1(t)-\lambda hP_1(t)+\lambda hP_0(t)+0(h).$$

$$\therefore\quad \frac{P_1(t+h)-P_1(t)}{h}=\lambda\{P_0(t)-P_1(t)\}+\frac{0(h)}{h}.$$

Hence proceeding to the limit $h\to 0+$, we get

$$P_1'(t)=\lambda\{P_0(t)-P_1(t)\}. \tag{5.11.2}$$

Case III. $i=0$.

The event 'no change in $(0, t+h)$' happens if and only if the events 'no change in $(0, t)$' and 'no change in $(t, t+h)$' both happen. As before E and E' being independent,

$$P_0(t+h)=P_0(t)\{1-\lambda h-0(h)-0(h)\}=P_0(t)-\lambda hP_0(t)+0(h).$$

$$\therefore\quad \frac{P_0(t+h)-P_0(t)}{h}=-\lambda P_0(t)+\frac{0(h)}{h}.$$

Proceeding to the limit $h \to 0+$, we get

$$P_0'(t) = -\lambda P_0(t). \tag{5.11.3}$$

From (5.11.1), (5.11.2) and (5.11.3) we get

$$\left.\begin{aligned} P_i'(t) &= \lambda P_{i-1}(t) - \lambda P_i(t) \text{ if } i \geq 1 \\ \text{and} \quad P_0'(t) &= -\lambda P_0(t). \end{aligned}\right\} \tag{5.11.4}$$

Solving (5.11.3), $P_0(t) = Ae^{-\lambda t}$, where A is an arbitrary constant.

We can assume the following initial conditions:

$$P_0(0) = 1 \text{ and } P_i(0) = 0 \text{ for all } i \geq 1. \tag{5.11.5}$$

Using $P_0(0) = 1$, we get $A = 1$.

$$\therefore \quad P_0(t) = e^{-\lambda t}. \tag{5.11.6}$$

Again, from (5.11.4), for $i = 1$,

$$P_1'(t) + \lambda P_1(t) = \lambda P_0(t)$$

or, $P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}$ by (5.11.6).

Multiplying by the integrating factor $e^{\lambda t}$ and integrating, we get

$$e^{\lambda t} P_1(t) = \lambda t + \text{constant}.$$

Using (5.11.5), $P_1(0) = 0$ and we get

$$P_1(t) = \lambda t e^{-\lambda t}. \tag{5.11.7}$$

Again, from (5.11.4), for $i = 2$,

$$P_2'(t) = \lambda P_1(t) - \lambda P_2(t)$$

or, $P_2'(t) + \lambda P_2(t) = \lambda^2 t e^{-\lambda t}$ by (5.11.7).

Multiplying by the integrating factor $e^{\lambda t}$ and integrating,

wet get $e^{\lambda t} P_2(t) = \dfrac{(\lambda t)^2}{2} + \text{constant}.$

Then using (5.11.5), $P_2(0) = 0$ and so

$$P_2(t) = \frac{(\lambda t)^2}{2!} e^{-\lambda t}.$$

Finally, by induction,

$$P_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \; i = 0, 1, 2, \ldots \ldots \tag{5.11.8}$$

This shows that the distribution of $X(t)$ is a Poisson distribution with parameter λt.

Remark: The *Poisson process* is a random phenomenon of counting numbers in a certain interval of time and where the three postulates (*i*), (*ii*), (*iii*) stated above are satisfied. In Practical field such counting may be the number of telephone calls or the number of radioactive emissions by a radioactive substance or the number of defective components of a certain instrument or the number of meteorite collisions, etc. In all such cases of counting process, Poisson distribution serves the appropriate model. Here λ can be interpreted as the average number of changes per unit time in the stochastic sense.

5.12. Transformation of Continuous Random Variable.

Let $y = g(x)$ be a given real valued continuous function defined for all real values of x and let X be a random variable defined on a given event space S. Then $g(X)$ is also a random variable defined by $[g(X)]\,\omega = g[X(\omega)]$, for all $\omega \in S$. If we denote $g(X)$ by Y, we say that $Y = g(X)$ determines a transformation of the random variable X.

THEOREM 5.12.1. *Let X be a continuous random variable and let $f_X(x)$ be the corresponding probability density function. Also let $y = g(x)$ be a continuously differentiable function for all values of x. If $f_Y(y)$ be the probability density function of the random variable Y, given by $Y = g(X)$ and if $\frac{dy}{dx}$ is either positive or negative for all x, then*

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|, \text{ where } y \in \text{range of } g\,. \tag{5.12.1}$$

and where we assume that $f_X(x)$ is defined for all values of x.

Proof: Let X be a continuous random variable. Let the random variable Y be defined by $Y = g(X)$, where $y = g(x)$ is a continuously differentiable function. We assume that $\frac{dy}{dx}$ is either positive or negative for all x.

Case I. $\frac{dy}{dx} > 0$ for all x. Here $y = g(x)$ is a strictly monotonically increasing function of x. Then the event $(X \leqslant x)$ implies and is implied by the event $(g(X) \leqslant g(x))$, i.e., $(X \leqslant x)$ and $(g(X) \leqslant g(x))$ represent the same event.

Hence $P(X \leqslant x) = P(g(X) \leqslant g(x))$

$= P(-\infty < Y \leqslant y)$ if $y = g(x)$ i.e.

or, $P(-\infty < X \leqslant x) = P(-\infty < Y \leqslant y)$ if $y \in$ range of g. (5.12.2)

or, $F_X(x) = F_Y(y)$.

where $F_X(x)$ and $F_Y(y)$ are the distribution functions of X and Y respectively.

The relation (5.12.2) shows that X being a continuous random variable and $y = g(x)$ being continuously differentiable function, Y is also a continuous random variable.

Differentiating (5.12.2) we get

$$\frac{d}{dx}\left\{F_X(x)\right\} = \frac{d}{dy}\left\{F_Y(y)\right\}\frac{dy}{dx}.$$

$$\text{or, } f_X(x) = f_Y(y)\frac{dy}{dx}.$$

$$\therefore \quad f_Y(y) = f_X(x)\frac{dx}{dy} = f_X(x)\left|\frac{dx}{dy}\right|, \text{ as } \frac{dy}{dx} > 0.$$

Case II. Let $\dfrac{dy}{dx} < 0$ for all x.

Here the function $y = g(x)$ is strictly monotonically decreasing, so that the event $(X \leqslant x)$ implies and is implied by the event $(g(X) \geqslant g(x))$. So here $(X < x)$ and $(g(X) > g(x))$ represent the same event.

$$\therefore \quad P(X < x) = P(g(X) > g(x)) = P(Y > y) = 1 - P(Y \leqslant y) \qquad (5.12.3)$$

or, $P(X \leqslant x) = 1 - P(Y \leqslant y)$, if $y = g(x)$

X being a continuous random variable. $P(X = x) = 0$.

$$\therefore \quad P(-\infty < X \leqslant x) = 1 - P(-\infty < Y \leqslant y)$$

or, $F_X(x) = 1 - F_Y(y)$, which indicates that Y is also continuous

$$\text{So, } \frac{d}{dx}\{F_X(x)\} = -\frac{d}{dy}\{F_Y(y)\}\frac{dy}{dx}$$

$$\text{or, } f_X(x) = -f_Y(y)\frac{dy}{dx}$$

$$\text{or, } f_Y(y) = -f_X(x)\frac{dx}{dy} = f_X(x)\left|\frac{dx}{dy}\right|, \text{ since } \frac{dx}{dy} < 0.$$

Hence the required result is established.

Here we note that in any case $f_Y(y) = 0$ if $y \notin$ range of g.

Remarks: 1. If $f_X(x)$ is undefined for some values of x, then 5.12.1 is true for those values of x, where $f_X(x)$ is defined.

2. If X be a discrete random variable and if the transformation $Y = g(X)$ be such that the corresponding real valued function $y = g(x)$ is continuous and strictly monotonic for all x, then the probability distribution of Y can be determined as follows:

Let x_i $(i = 0, \pm 1, \pm 2, \ldots\ldots)$ be the spectrum of X. Then the spectrum of Y is given by $y_i = g(x_i)$; $i = 0, \pm 1, \pm 2, \ldots\ldots$ (5.12.4). Since the transformation $Y = g(X)$ is strictly monotonic, the event $(X = x_i)$ implies and is implied by

$\{g(X) = g(x_i)\}$ i.e. $(Y = y_i)$. So,

$$P(X = x_i) = P\{g(X) = g(x_i)\} = P(Y = y_i),$$

i.e., $f_{y_i} = f_{x_i}$, for $i = 0, \pm 1, \pm 2, \ldots\ldots$ (5.12.5).

3. Since the inverse function $x = g^{-1}(y)$ exists under the condition of the theorem 5.12.1, the formula (5.12.1) can be expressed as $f_Y(y) = f_X\{g^{-1}(y)\}\left|\dfrac{dx}{dy}\right|$. (5.12.6)

4. If the function $g(x)$ in theorem 5.12.1 is not strictly monotonic, the formula (5.12.1) cannot be applied to determine the distribution of Y. In this case, it is possible to find the distribution of Y by applying probability differential or other suitable methods.

5.13. Mixed Distribution.

So far we have discussed distributions which are either discrete or continuous. But there are distributions which are neither discrete nor continuous. In fact, there are probability distributions where the corresponding distribution is partly discrete and partly continuous. Such a distribution is called a mixed distribution. We give below a formal definition of a mixed distribution.

A distribution is called a *mixed distribution*, if the corresponding distribution function $F(x)$ can be expressed as a convex combination of the form

$$F(x) = cF_1(x) + (1 - c)F_2(x) \qquad (5.13.1)$$

where $F_1(x)$ is the distribution function of a discrete random variable and $F_2(x)$ is that of a continuous random variable and $0 < c < 1$.

We complete the discussion by giving the following example of a mixed distribution.

Let X be a random variable with distribution function $F(x)$ given by

$$F(x)=\begin{cases} 0, & x<0 \\ \dfrac{x+1}{2}, & 0\leqslant x<1 \\ 1, & 1\leqslant x. \end{cases}$$

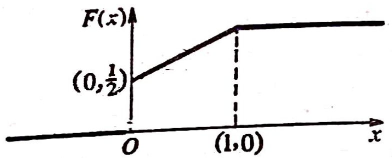


Fig. 5.13.1 Distribution curve of a mixed distribution.

From fig. 5.13.1, we see that $F(x)$ has a jump discontinuity at $x=0$. In fact, $F(x)$ is not always continuous, nor is it a step function. Accordingly, the corresponding distribution is a mixed distribution.

We can write

$$F(x)=\tfrac{1}{2}F_1(x)+\tfrac{1}{2}F_2(x),$$

where $$F_1(x)=\begin{cases} 0, & x<0 \\ 1, & x\geqslant 0 \end{cases}$$

and $$F_2(x)=\begin{cases} 0, & x\leqslant 0 \\ x, & 0<x<1 \\ 1, & x\geqslant 1. \end{cases}$$

$F_1(x)$ and $F_2(x)$ are the distribution functions of a discrete and a continuous distribution respectively. The probability density function $f(x)$ corresponding to $F_2(x)$ is given by

$$f(x)=\begin{cases} 1, & 0<x<1 \\ 0, & \text{if } -\infty<x<0, \text{ and } 1<x<\infty. \end{cases}$$

We observe that $f(x)$ is undefined at $x=0$ and at $x=1$.

In fact $L\,F_2'(0)=0$, $R\,F_2'(0)=1$, etc.

5·14. Illustrative Examples.

Ex. 1. *Five balls are drawn from an urn containing 4 white and 6 black balls. Find the probability distribution of the number of white balls drawn without replacement.*

Let X be the random variable denoting the number of white balls drawn from the urn. Then the spectrum of X is the set $\{0, 1, 2, 3, 4\}$.

Now $P(X=0)=\dfrac{{}^6P_5}{{}^{10}P_5}=\dfrac{1}{42}$, $P(X=1)=\dfrac{{}^4C_1\times{}^6C_4\times\lfloor 5}{{}^{10}P_5}=\dfrac{5}{21}$,

$$P(X=2)=\frac{{}^4C_2\times{}^6C_3\times\lfloor 5}{{}^{10}P_5}=\frac{10}{21},$$

$$P(X=3)=\frac{{}^4C_3\times{}^6C_2\times\lfloor 5}{{}^{10}P_5}=\frac{5}{21},$$

$$P(X=4)=\frac{{}^4C_4\times{}^6C_1\times\lfloor 5}{{}^{10}P_5}=\frac{1}{42}.$$

Hence the required distribution at X is given by the spectrum $\{0, 1, 2, 3, 4\}$ with

$P(X=0)=\frac{1}{42}$, $P(X=1)=\frac{5}{21}$, $P(X=2)=\frac{10}{21}$, $P(X=3)=\frac{5}{21}$, and $P(X=4)=\frac{1}{42}$.

Ex. 2. *Consider the random experiment of tossing a fair coin till a head appears for the first time. Let X be the number of tosses required. Find the distribution of X.*

The spectrum of X is the set $\{1, 2, 3, \ldots\ldots\}$.

Now $P(X=1)=P\{H\}=\frac{1}{2}$, $P(X=2)=P\{(T, H)\}=(\frac{1}{2})^2$,

$P(X=3)=P\{(T, T, H)\}=(\frac{1}{2})^3, \ldots\ldots,$

$$P(X=n)=P\{(\underbrace{T, T, \ldots\ldots T}_{(n-1)\text{ times}}, H)\}=(\tfrac{1}{2})^n$$

and so on, where 'H' denotes the outcome head in the 'first toss', (T, H) denotes the outcome 'Tail in the first toss and head in the second toss', etc. Thus the required distribution is given by $X=i$, $i=1, 2, \ldots\ldots$, with

$$P(X=i)=(\tfrac{1}{2})^i.$$

Ex. 3. *Show that the function* $|x|$ *in* $(-1, 1)$ *and zero elsewhere is a possible density function, and find the corresponding distribution function.*

Let $f(x)=\begin{cases} |x|, & -1<x<1 \\ 0, & \text{elsewhere.}\end{cases}$

We see that $f(x) \geqslant 0$ for every x.

Also, $\int_{-\infty}^{\infty} f(x)\,dx=\int_{-\infty}^{-1} f(x)\,dx+\int_{-1}^{1} f(x)\,dx+\int_{1}^{\infty} f(x)\,dx$

$$=0+\int_{-1}^{1}|x|\,dx+0$$

$=2\int_0^1 x\,dx$, since $|x|$ is an even function and $|x|=x$ for every $x\in(0,1)$

$=1.$

Hence, $f(x)$ is a possible probability density function of some distribution.

Now let $F(x)$ be the corresponding distribution function.

If $-\infty<x\leqslant-1$, $F(x)=\int_{-\infty}^{x} f(t)\,dt=0.$

If $-1<x<0$, $F(x)=\int_{-\infty}^{x} f(t)\,dt=\int_{-\infty}^{-1} f(t)\,dt+\int_{-1}^{x} f(t)\,dt$

$$=0+\int_{-1}^{x} f(t)\,dt=\int_{-1}^{x}(-t)\,dt=\frac{1}{2}-\frac{x^2}{2}.$$

If $0\leqslant x<1$, $F(t)=\int_{-\infty}^{x} f(t)\,dt$

$$=\int_{-\infty}^{-1} f(t)\,dt+\int_{-1}^{0} f(t)\,dt+\int_{0}^{x} f(t)\,dt$$

$$=0-\int_{-1}^{0} t\,dt+\int_{0}^{x} t\,dt$$

$$=\frac{1}{2}+\frac{x^2}{2}.$$

If $1\leqslant x<\infty$, $F(x)=\int_{-\infty}^{x} f(t)\,dt$

$$=\int_{-\infty}^{-1} f(t)\,dt+\int_{-1}^{1} f(t)\,dt+\int_{1}^{x} f(t)\,dt$$

$$=0+1+0=1.$$

Hence the distribution function is given by

$$F(x)=\begin{cases} 0, & -\infty<x\leqslant-1 \\ \frac{1}{2}-\frac{x^2}{2}, & -1<x<0 \\ \frac{1}{2}+\frac{x^2}{2}, & 0\leqslant x<1 \\ 1, & 1\leqslant x<\infty. \end{cases}$$

Ex. 4. *Can the following be probability mass functions?*

(a) $f(x)=\begin{cases} 2 & \textit{for } x=\frac{1}{2} \\ 1 & \textit{for } x=\frac{1}{4} \\ -1 & \textit{for } x=\frac{3}{4} \\ 0, & \textit{elsewhere.}\end{cases}$

(b) $f(x)=\begin{cases} \frac{1}{8} & \textit{for } x=1 \\ \frac{2}{8} & \textit{for } x=2 \\ \frac{3}{8} & \textit{for } x=3 \\ 0, & \textit{elsewhere.}\end{cases}$

(c) $f(x)=\begin{cases} 0{\cdot}1 & \textit{for } x=-5 \\ 0{\cdot}5 & \textit{for } x=-1 \\ 0{\cdot}2 & \textit{for } x=0 \\ 0{\cdot}2 & \textit{for } x=1 \\ 0, & \textit{elsewhere.}\end{cases}$

(a) Since $f(\frac{3}{4})=-1<0$, $f(x)$ is not a probability mass function.

(b) Although $f(x)\geqslant 0$ for every mass point, $\sum_x f(x)\neq 1$.

Hence $f(x)$ is not a probability mass function.

(c) $f(x)\geqslant 0$ for every spectrum point and $\sum_x f(x)=1$, hence $f(x)$ is a probability mass function of a distribution.

Ex. 5. *Evaluate the distribution function of the following distribution: Spectrum of the random variable* X *is* $\{-1, 0, 2, 3\}$ *with*

$$P(X=-1)=\tfrac{1}{7},\ P(X=0)=\tfrac{2}{7},\ P(X=2)=\tfrac{3}{7},\ P(X=3)=\tfrac{1}{7}.$$

Let $F(x)$ be the distribution function.

If $-\infty<x<-1$, $F(x)=0.$

If $-1\leqslant x<0$, $F(x)=P(X=-1)=\frac{1}{7}.$

If $0\leqslant x<2$, $F(x)=P(X=-1)+P(X=0)=\frac{1}{7}+\frac{2}{7}=\frac{3}{7}.$

If $2 \leqslant x < 3$, $F(x)=P(X=-1)+P(X=0)+P(X=2)$

$$=\tfrac{1}{7}+\tfrac{2}{7}+\tfrac{3}{7}=\tfrac{6}{7}.$$

If $3 \leqslant x<\infty$, $F(x)=P(X=-1)+P(X=0)+P(X=2)+P(X=3)$

$$=\tfrac{1}{7}+\tfrac{2}{7}+\tfrac{3}{7}+\tfrac{1}{7}$$
$$=1.$$

Ex. 6. *Let $F(x)$ be the distribution function of a random variable X. Prove that*

(*i*) $P(a \leqslant X < b)=F(b-0)-F(a-0)$,

(*ii*) $P(a \leqslant X \leqslant b)=F(b)-F(a-0)$.

(*i*) The event $(a \leq X<b)$ can be expressed as

$(a<X <b)+(X=a)$.

$\therefore\ P(a \leqslant X < b)=P(a < X < b)+P(X=a)$, (5.14.1)

where we note that $(a < X < b)$, $(X=a)$ are mutually exclusive events.

Again we can write

$$(a < X \leqslant b)=(a < X < b)+(X=b).$$

So $P(a < X \leq b)=P(a < X < b)+P(X=b)$.

$$P(a < X < b)=P(a <X \leqslant b)-P(X=b).$$

Hence, by (5.14.1) we get

$$\begin{aligned}P(a \leqslant X < b)&=P(a < X \leqslant b)-P(X=b)+P(X=a)\\&=F(b)-F(a)-F(b)+F(b-0)+F(a)-F(a-0)\\&=F(b-0)-F(a-0).\end{aligned}$$

(*ii*) We have

$$(a \leqslant X \leq b)=(a < X \leq b)+(X=a),$$

where $(a < X \leq b)$, $(X=a)$ are two mutually exclusive events.

So $P(a \leqslant X \leqslant b)=P(a < X \leq b)+P(X=a)$

$$=F(b)-F(a)+F(a)-F(a-0)$$
$$=F(b)-F(a-0).$$

Ex. 7. *Can the following function be a distribution function?*

$$F(x)=\begin{cases}0, & -\infty < x < 0\\ \tfrac{1}{5}, & 0 \leqslant x < 1\\ \tfrac{3}{5}, & 1 \leqslant x < 3\\ 1, & 3 \leqslant x < \infty.\end{cases}$$

If so, find the spectrum and probability mass function.

It is clear that $F(x)$ is monotonically non-decreasing and non-negative and $F(\infty)=1$, $F(-\infty)=0$. $F(x)$ is a step function, discontinuous to the left of the three step points 0, 1, 3 and continuous to the right everywhere. Hence, $F(x)$ is a possible distribution function of a discrete random variable X. The spectrum of

X is $\{0, 1, 3\}$ with

$$P(X=0)=F(0)-F(0-0)=\tfrac{1}{5},$$
$$P(X=1)=F(1)-F(1-0)=\tfrac{3}{5}-\tfrac{1}{5}=\tfrac{2}{5},$$
$$P(X=3)=F(3)-F(3-0)=1-\tfrac{3}{5}=\tfrac{2}{5},$$

which give the probability masses at the spectrum points and these probability masses determine the required probability mass function.

Ex. 8. *Determine the value of the constant C such that $f(x)$ defined by*

$$f(x)=\begin{cases}Cx(1-x), & 0 < x < 1\\ 0, & \textit{elsewhere}\end{cases}$$

is a probability density function. Find the corresponding distribution function and $P(X > \tfrac{1}{2})$.

In order that $f(x)$ is a possible probability density function, we must have

$$\int_{-\infty}^{\infty} f(x)\,dx=1$$

i.e., $C \int_0^1 x(1-x)\,dx=1$

i.e., $C=6$.

Let $F(x)$ be the corresponding distribution function.

In $-\infty < x < 0$, $F(x)=0$,

in $0 \leq x \leq 1$, $F(x)=6\int_0^x t(1-t)\,dt=3x^2-2x^3$,

in $1 < x <\infty$, $F(x)=6\int_0^1 t(1-t)dt=1$.

$$\therefore\ F(x)=\begin{cases}0, & -\infty < x < 0\\ 3x^2-2x^3, & 0 \leqslant x \leqslant 1\\ 1, & 1 < x < \infty.\end{cases}$$

Finally, $P(X \leq \frac{1}{3}) = F(\frac{1}{3}) = 3(\frac{1}{3})^2 - 2(\frac{1}{3})^3 = \frac{7}{27}$

$\therefore \quad P(X > \frac{1}{3}) = \frac{20}{27}$

Ex. 9. *Distribution function $F(x)$ of a random variable X is given by*

$$F(x) = \begin{cases} 1 - \frac{1}{2}e^{-x}, & x \geq 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find $P(X=0)$ and $P(X>1)$.

$$P(X=0) = F(0) - F(0-0) = \tfrac{1}{2} - 0 = \tfrac{1}{2}.$$

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = \frac{1}{2e}.$$

Ex. 10. *Let X be a random variable such that*

$$P(a < X \leq b) = e^{-\lambda a} - e^{-\lambda b} \text{ and } P(X \leq 0) = 0,$$

where $0 \leq a < b < \infty$ *and* λ *is a suitable positive real number.*

Find the distribution of X.

Let $F(x)$ be the corresponding distribution function.

In $-\infty < x \leq 0$, $F(x) = 0$.

In $0 < x < \infty$, $F(x) = P(0 < X \leq x) = 1 - e^{-\lambda x}$

$$\therefore \quad F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & 0 < x < \infty. \end{cases}$$

Ex. 11. *The distribution function of a random variable* X *is given by*

$$F(x) = \begin{cases} 0, & x \leq 0 \\ x, & 0 < x \leq 1 \\ 1, & 1 < x. \end{cases}$$

Find the probability density function and evaluate

$$P(\cdot 3 < X \leq \cdot 5).$$

Let $f(x)$ be the probability density function.

Then

$$f(x) = F'(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & -\infty < x < 0,\ 1 < x < \infty. \end{cases}$$

We observe that $f(x)$ is undefined at $x=0$ and at $x=1$.

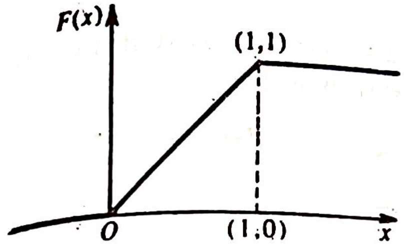


Fig. 5.14.1 Distribution curve

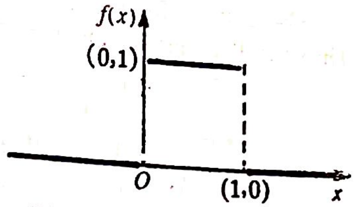


Fig. 5.14.2 Density curve

$$P(\cdot 3 < X \leq \cdot 5) = F(\cdot 5) - F(\cdot 3) = \cdot 2.$$

Ex. 12. *The spectrum of the random variable X consists of the points* 1, 2, ..., *n and* $P(X=i)$ *is proportional to* $\frac{1}{i(i+1)}$. *Determine the distribution function of* X. *Compute* $3 < X \leq n$ *and* $P(X > 5)$.

We have, $f_i = P(X=i) \propto \frac{1}{i(i+1)}$.

Let $f_i = K \cdot \frac{1}{i(i+1)}$, where K is a positive constant to be determined from $\sum_{i=1}^{n} f_i = K \sum_{i=1}^{n} \frac{1}{i(i+1)} = 1$.

$$\therefore \quad K \sum_{i=1}^{n} \left(\frac{1}{i} - \frac{1}{i+1}\right) = 1$$

or, $K\left(1 - \frac{1}{n+1}\right) = 1$, i.e., $K = \frac{n+1}{n}$.

Hence, $f_i = P(X=i) = \frac{n+1}{n} \frac{1}{i(i+1)}$.

Let $F(x)$ be the distribution function of X. Then

$F(x) = P(-\infty < X \leq x)$, where $i \leq x < i+1$, $i = 1, 2, \ldots n-1$

$$= \sum_{r=1}^{i} P(X=r)$$

$$= \frac{n+1}{n} \sum_{r=1}^{i} \frac{1}{r(r+1)}$$

$$= \frac{n+1}{n}\left(1 - \frac{1}{i+1}\right) = \frac{n+1}{n} \frac{i}{i+1}.$$

Thus $F(x)=\frac{n+1}{n}\frac{i}{i+1}$; $i\leqslant x<i+1$ for $i=1, 2,\ldots n-1$.

Also $F(x)=1$ for all $x\geqslant n$, and $F(x)=P(X\leqslant x)=0$ if $x<1$.

Thus the distribution function $F(x)$ is given by

$$F(x)=\begin{cases}\frac{n+1}{n}\frac{i}{i+1}, & \text{if } i\leqslant x<i+1, i=1, 2,\ldots\ldots, n-1\\ 1 & , \text{if } x\geqslant n\\ 0 & , \text{if } x<1.\end{cases}$$

Now $P(3<X\leq n)=F(n)-F(3)=1-\frac{n+1}{n}\frac{3}{4}=\frac{n-3}{4n}$.

$$P(X>5)=1-P(X\leq 5)=1-F(5)=1-\frac{n+1}{n}\cdot\frac{5}{6}=\frac{n-5}{6n}.$$

Ex. 13. *Consider the distribution function of X given by*

$$F(x)=\begin{cases}0, \text{ for } x<0\\ 1-\frac{1}{2}e^{-x} \text{ for } x\geq 0.\end{cases}$$

Determine $P(X=0)$ and $P(X>0)$. [C.H. (Math.) '81]

$$P(X=0)=F(0)-F(0-0)=(1-\tfrac{1}{2})-0=\tfrac{1}{2}.$$

$$P(X>0)=1-P(-\infty<X\leqslant 0)=1-F(0)=1-(1-\tfrac{1}{2})=\tfrac{1}{2}.$$

Ex. 14. *A motorist encounters n consecutive traffic lights, each likely to be red with probability p or green with probability $q=1-p$. Let $x-1$ be the number of green lights passed by the motorist before being stopped first time by a red light. Show that the probability distribution of x is given by*

$$f(x)=q^{x-1}p, x=1, 2,\ldots\ldots, n,$$
$$=q^n, x=n+1.$$

[C. H. (Math.) '80]

Let 'success' S denote encountering a red light and 'failure' F denote encountering a green light. Then the event space consists of finite sequence of outcomes $S, FS, FFS, FFFS,\ldots\ldots, \underbrace{FF\ldots\ldots F}_{n \text{ times}}$.

The corresponding random variable X can take values $1, 2, 3,\ldots\ldots, n, n+1$, where $X-1$ is the number of failures before the first success.

$\therefore$ $P(X=1)=p$, $P(X=2)=qp$, $P(X=3)=q^2p,\ldots\ldots,$
$P(X=n)=q^{n-1}p$, $P(X=n+1)=q^n$.

Hence, the probability distributian of X is given by

$$f(x)=P(X=x)=q^{x-1}p, \text{ for } x=1, 2,\ldots\ldots n,$$
$$=q^n \text{ for } x=n+1.$$

Ex. 15. *The life in hours of a certain kind of radio tube has the probability density function*

$$f(x)=\begin{cases}\frac{100}{x^2}, & x\geq 100\\ 0, & \textit{elsewhere}.\end{cases}$$

What is the probability that none of three tubes in a given radio set will have to be replaced during the first 150 hours of operation? What is the probability that all three of the original tubes will have to be replaced during the first 150 hours?

Let A_i denote the event that 'the life of the ith tube is at least 150 hours' $(i=1, 2, 3)$.

$$\text{Then } P(A_i)=\int_{150}^{\infty}\frac{100}{x^2}\,dx=\tfrac{2}{3}.$$

Now the events A_1, A_2 and A_3 are independent. Hence the probability that all the three tubes last 150 hours or more (so that none of them will have to be replaced during the first 150 hours of operation)

$$=P(A_1A_2A_3)=P(A_1)P(A_2)P(A_3)=\tfrac{8}{27}.$$

Again $\overline{A}_i$ is the event that 'the life of the tube is less than 150 hours'. $P(\overline{A}_i)=1-P(A_i)=\frac{1}{3}$.

Hence, the probability that all three tubes will have to be replaced during the first 150 hours

$$=P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)=\tfrac{1}{27}.$$

Ex. 16. *The probability that a screw manufactured by a machine to be defective is $\frac{1}{50}$. A lot of 6 screws are taken at random. Find the probability that (i) there are exactly 2 defective screws in the lot, (ii) no defective screw and (iii) at most 2 defective screws.*

We consider the event 'getting a screw defective' as 'success'. Then p=probability of success$=\frac{1}{50}$, $n=6$. So if X be the random variable corresponding to the number of defective screws, then X is a binomial $(6, p)$ variate.

(i) The required probability$=P(X=2)=\binom{6}{2}(\frac{1}{50})^2(\frac{49}{50})^4$.

(ii) The required probability$=P(X=0)=\binom{6}{0}(\frac{1}{50})^0(\frac{49}{50})^6=(\frac{49}{50})^6$.

(iii) The required probability$=P(X=0)+P(X=1)+P(X=2)$

$$=(\tfrac{49}{50})^6+\binom{6}{1}(\tfrac{1}{50})(\tfrac{49}{50})^5+\binom{6}{2}(\tfrac{1}{50})^2(\tfrac{49}{50})^4.$$

Ex. 17. *The probability of a product produced by a machine to be defective is 0·01. If 30 products are taken at random, find the probability that exactly 2 will be defective. Approximate by Poisson distribution and evaluate the error in the approximation.*

As in Ex. 16, required probability$=\binom{30}{2}(\cdot01)^2(\cdot99)^{28}=0\cdot0276$.

Since the probability of success is small, we approximate by Poisson distribution, the parameter of the distribution being $\mu=np=30\times\cdot01=0\cdot3$.

Hence the probability of getting exactly 2 defective

$$=\frac{\mu^2}{2!}e^{-\mu}=\frac{(0\cdot3)^2}{2!}e^{-0\cdot3}=0\cdot03337.$$

$\therefore$ the error in the approximation$=0\cdot03337-0\cdot03276=0\cdot00061$.

Ex. 18. *Some airlines find that each passenger who reserves a seat fails to turn up with probability 0·1 independently of other passengers of these airlines. Airlline A always sells 10 tickets for their 9-seat aeroplane while airline B always sells 20 tickets for their 18-seat aeroplane. Using Poisson approximation to binomial distribution, find which one of A and B is more often overbooked.*

[C. H. (Math.) '83]

Let X and Y be the random variables corresponding to the number of passengers who failed to turn up in the airlines A and B respectively.

For the random variable X, we have $n=10$, $p=0\cdot1$ and so $\mu=1$. X is approximately a Poisson-variate with parameter 1. Similarly Y is approximately a Poisson-variate with parameter 2 ($n=20$, $p=0\cdot1$, $\mu=2$).

The probability that the airline A is overbooked

$$=P(X=0)=\frac{e^{-1}.1}{0!}=e^{-1}.$$

Again the probability that the airline B is overbooked

$$=P(Y=0)+P(Y=1)$$

$$=\frac{e^{-2}2^0}{0!}+\frac{e^{-2}.2^1}{1!}=\frac{3}{e^2}.$$

Now $\frac{3}{e^2}-\frac{1}{e}=\frac{3-e}{e^2}>0$, since $2<e<3$.

$\therefore$ $\frac{3}{e^2}>\frac{1}{e}$ and hence the airline B is more often overbooked.

Ex. 19. *If there is a war every 15 years on the average, then find the probability that there will be no war in 25 years.*

λ=number of changes per unit of time on the average$=\frac{1}{15}$. Let X be the random variable denoting the number of wars in the interval (0, 25), when the unit of time is one year, then X is Poisson distributed with parameter $\mu=\lambda t=\frac{1}{15}\times25=\frac{5}{3}$.

$\therefore$ probability of no war in the given interval of time

$$=P(X=0)=\frac{e^{-\mu}\mu^0}{0!}=e^{-\frac{5}{3}}.$$

Ex. 20. *A radio active source emits on the average 2·5 particles per second. Calculate the probability that 3 or more particles will be emitted in an interval of 4 seconds.* [C.H. (Math.) '87]

λ=number of changes per unit time on the average $=2\cdot5$.

Also the given interval is (0, 4), where the unit of time is one second. If X be the random variable denoting the number of particles emitted in the given interval, then X is Poisson-distributed with parameter $\mu=\lambda t=10$

$\therefore$ the required probability$=P(X\geqslant3)=1-P(X<3)$

$$=1-\{P(X=0)+P(X=1)+P(X=2)\}$$

$$=1-e^{-10}-10\,e^{-10}-\tfrac{100}{2}e^{-10}=1-61\,e^{-10}.$$

Ex. 21. *A car-hire firm has two cars, which it hires out by the day. The number of demands for a car on each day is Poisson distributed with parameter 1·5. Calculate the proportion of days*

on which neither of the cars is used, and the proportion of days on which some demand cannot be met for lack of cars.

[*C.H. (Math.)* '80]

$(e^{-1.5}=0.223)$

Let X be the random variable denoting the number of demands for a car on any day. Then X is Poisson distributed with parameter 1·5

(*a*) Proportion of days on which neither car is used

$$=P(X=0)=e^{-1.5}=0.223.$$

(*b*) Proportion of days on which some demand is refused

$$=P(X>2)=1-P(X\leqslant 2)$$

$$=1-\{P(X=0)+P(X=1)+P(X=2)\}$$

$$=1-e^{-1.5}\left\{1+1.5+\frac{(1.5)^2}{2}\right\}$$

$$=0.1916.$$

Ex. 22. *Let X denote the tangent of an angle (measured in radians) chosen at random from $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$. Find the distribution of X.*

Let H be the random variable denoting the angle chosen. Then H is uniformly distributed whose distribution function $F(\theta)$ is given by

$$F(\theta)=\begin{cases}0, & -\infty<\theta<-\frac{\pi}{2}\\ \dfrac{\theta-(-\frac{\pi}{2})}{\pi}, & -\frac{\pi}{2}\leqslant\theta<\frac{\pi}{2}\\ 1, & \frac{\pi}{2}\leqslant\theta<\infty.\end{cases}$$

Now $X=\tan H$, and hence

$$P(X\leqslant x)=P(\tan H\leqslant x)=P\left(-\frac{\pi}{2}<H\leqslant\tan^{-1}x\right)$$

$$=\frac{\tan^{-1}x-\left(-\frac{\pi}{2}\right)}{\pi}=\frac{1}{2}+\frac{1}{\pi}\tan^{-1}x,-\infty<x<\infty.$$

Thus the distribution function of X is given by

$$F_X(x)=\frac{1}{2}+\frac{1}{\pi}\tan^{-1}x,-\infty<x<\infty.$$

Ex. 23. *A point P is chosen at random on a circle of radius a and A is a fixed point on the circle. Find the probability that the chord AP will exceed the length of an equilateral triangle inscribed in the circle.*

Let X be the random variable corresponding to the angle $A'AP$, where AA' is the diameter through A. The distribution of X is given by the density function

$$f(x)=\begin{cases}\frac{1}{\pi}, & -\frac{\pi}{2}\leqslant x\leqslant\frac{\pi}{2}\\ 0, & \text{elsewhere.}\end{cases}$$

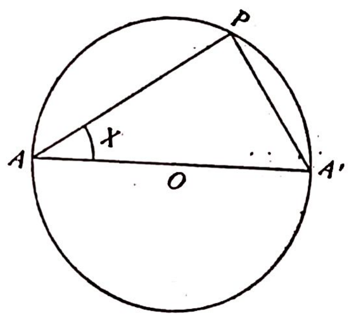


Fig. 5.14.3

The side of an equilateral triangle inscribed within a circle being $\sqrt{3}a$, where a is the radius of the circle, the required probability

$$=P(2a\cos X>\sqrt{3}a)=P\left(\cos X>\frac{\sqrt{3}}{2}\right)$$

$$=P\left(|X|<\frac{\pi}{6}\right)=P\left(-\frac{\pi}{6}<X<\frac{\pi}{6}\right)=\int_{-\frac{\pi}{6}}^{\frac{\pi}{6}}f(x)\,dx=\tfrac{1}{3}.$$

Ex. 24. *Three concentric circles of radii $\frac{1}{\sqrt{3}}$, 1 and $\sqrt{3}$ feet are drawn on a target board. If a shot falls within the innermost*

circle 3 points are scored; if it falls within the next two rings, then the scores are respectively 2 and 1 and the score is zero if the shot falls outside the outermost circle. If the probability density of the distance of the hit from the centre of the target is $\frac{2}{\pi}\frac{1}{1+r^2}$, *then find the probability distribution of the score.*

Let X be the random variable corresponding to the score. Let R be the random variable denoting the distance of the hit P from the centre O of the target. Evidently R is a continuous random variable, whose density function is given as

$$f(r)=\frac{2}{\pi}\;\frac{1}{1+r^2},\; 0\leqslant r<\infty.$$

The spectrum of X is $\{0, 1, 2, 3\}$.

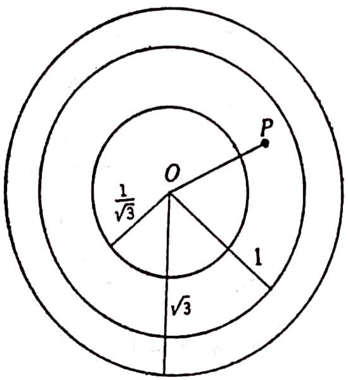


Fig. 5.14.4

Now the event '$X=0$' happens if and only if the event '$R>\sqrt{3}$' happens.

$$\therefore\quad P(X=0)=P(R>\sqrt{3})=1-P(0<R\leqslant\sqrt{3})$$

$$=1-\int_0^{\sqrt{3}}\frac{2}{\pi}\;\frac{dr}{1+r^2}=\frac{1}{3}.$$

Similarly $P(X=1)=P(1<R<\sqrt{3})=P(1<R\leqslant\sqrt{3})$

$$=\frac{2}{\pi}\int_1^{\sqrt{3}}\frac{dr}{1+r^2}=\frac{1}{6},$$

$$P(X=2)=P\left(\frac{1}{\sqrt{3}}<R<1\right)=P\left(\frac{1}{\sqrt{3}}<R\leqslant 1\right)$$

$$=\frac{2}{\pi}\int_{\frac{1}{\sqrt{3}}}^{1}\frac{dr}{1+r^2}=\frac{1}{6},$$

$$P(X=3)=P\left(0<R<\frac{1}{\sqrt{3}}\right)=P\left(0<R\leqslant\frac{1}{\sqrt{3}}\right)$$

$$=\frac{2}{\pi}\int_0^{\frac{1}{\sqrt{3}}}\frac{dr}{1+r^2}=\frac{1}{3}.$$

Hence the required probability distribution of the score X is given by, $X=i,\; i=0, 1, 2, 3$ with

$$P(X=0)=\tfrac{1}{3},\; P(X=1)=\tfrac{1}{6}=P(X=2),\; P(X=3)=\tfrac{1}{3}.$$

Ex. 26. *If X is normal (m, σ), then prove that*

$$P(a<X<b)=\Phi\left(\frac{b-m}{\sigma}\right)-\Phi\left(\frac{a-m}{\sigma}\right)$$

and $P(|X-m|>a\sigma)=2\{1-\Phi(a)\}$,

where $\Phi(x)$ denotes the standard normal distribution function.

$P(a<X<b)=P(a<X\leqslant b)$

$$=P\left(\frac{a-m}{\sigma}<\frac{X-m}{\sigma}\leqslant\frac{b-m}{\sigma}\right)=P\left(\frac{a-m}{\sigma}<Y\leqslant\frac{b-m}{\sigma}\right),$$

where $Y=\dfrac{X-m}{\sigma}$ is the standard normal variate.

$$=\Phi\left(\frac{b-m}{\sigma}\right)-\Phi\left(\frac{a-m}{\sigma}\right).$$

Again $P(|X-m|>a\sigma)=1-P(|X-m|\leqslant a\sigma)$

$$=1-P(-a\sigma<X-m\leqslant a\sigma)$$

$$=1-P(m-a\sigma<X\leqslant a\sigma+m)$$

$$=1-\Phi\left(\frac{a\sigma+m-m}{\sigma}\right)+\Phi\left(\frac{m-a\sigma-m}{\sigma}\right)\text{ by first part.}$$

$$=1-\Phi(a)+\Phi(-a).$$

Now $\Phi(-a)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-a} e^{-\frac{t^2}{2}}\,dt=\frac{1}{\sqrt{2\pi}}\underset{B\to-\infty}{Lt}\int_{B}^{-a} e^{-\frac{t^2}{2}}\,dt$

$=\frac{1}{\sqrt{2\pi}}\underset{B\to-\infty}{Lt}\int_{a}^{-B} e^{-\frac{u^2}{2}}\,du$, where $u=-t$

$=\frac{1}{\sqrt{2\pi}}\int_{a}^{\infty} e^{-\frac{u^2}{2}}\,du=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}\,du-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-\frac{u^2}{2}}\,du$

$=1-\Phi(a)$.

$\therefore\ P(|X-m|>a\sigma)=1-\Phi(a)+1-\Phi(a)=2\{1-\Phi(a)\}$.

Ex. 26. *A ray of light is sent in a random direction towards the x-axis from the station $Q(0,1)$ on the y-axis and the ray meets the x-axis at a point P. Find the probability function of the abscissae of P.* [C.H. (Math.) '82]

The line through $Q(0, 1)$ cuts the x-axis at $(-\tan\theta, 0)$, where θ varies from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$. The point P also moves along x-axis from $-\infty$ to ∞.

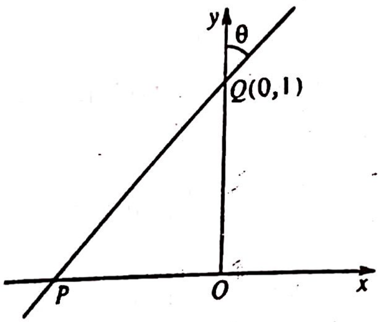


Fig. 5.14.5

The density function of H is $f(\theta)=\frac{1}{\pi}$, $-\frac{\pi}{2}\le\theta\le\frac{\pi}{2}$, where H is the random variable corresponding to θ. We now find the distribution of X, where X is the random variable defined by

$$X=-\tan H.$$

In real variable $x=-\tan\theta$.

$\therefore\ \frac{dx}{d\theta}=-\sec^2\theta<0$ for all $\theta\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$.

$\therefore$ the distribution of X is given by the density function $g(x)$, where

$$g(x)=f(\theta)\left|\frac{d\theta}{dx}\right|=\frac{1}{\pi}\cdot\frac{1}{1+x^2},\ -\infty<x<\infty.$$

This shows that X is Cauchy-distributed.

Ex. 27. *Let X be a standard normal variate. Find the probability density function of Y, where $Y=\frac{1}{2}X^2$.* [C.H. (Math.) '81, '84, '86, '93]

Here $Y=\frac{1}{2}X^2$.

In real variable, $y=\frac{1}{2}x^2$, $-\infty<x<\infty$.

$\therefore\ \frac{dy}{dx}=x$, which changes sign in $-\infty<x<\infty$.

Now we have

$$P(y<Y\le y+dy)=P(\tfrac{1}{2}x^2<\tfrac{1}{2}X^2\le\tfrac{1}{2}(x+dx)^2)$$
$$=P(x^2<X^2\le(x+dx)^2).$$

Let $x>0$. Then the event $(x^2<X^2\le(x+dx)^2)$ can be expressed as $(x<X\le x+dx)\cup(-(x+dx)\le X<-x)$, where we note that the two events $(x<X\le x+dx)$ and $(-(x+dx)\le X<-x)$ are mutually exclusive.

$\therefore\ P(x^2<X^2\le(x+dx)^2)$

$$=P(x<X\le x+dx)+P(-(x+dx)\le X<-x).$$

Now due to symmetry,

$$P(x<X\le x+dx)=P(-(x+dx)\le X<-x)$$

$\therefore\ P(y<Y\le y+dy)=2\,P(x<X\le x+dx)$

or, $f_Y(y)\,dy=2\,f_X(x)\,dx$,

where $f_X(x)$ and $f_Y(y)$ are the density functions of X and Y respectively. Now $f_X(x)=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}$.

$\therefore\ f_Y(y)=2\,f_X(x)\,\frac{dx}{dy}=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}\cdot\frac{2}{x}=\frac{e^{-y}y^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})}$, as $x=\sqrt{2y}$.

Thus if $x>0$, $f_Y(y)=\frac{e^{-y}y^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})}$, $0<y<\infty$.

It can be shown similarly that if $x < 0$,

$$f_Y(y) = \frac{e^{-y} y^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})},\ 0 < y < \infty.$$

Thus we get, in either case, $f_Y(y) = \dfrac{e^{-y} y^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})}$, $0 < y < \infty$, which shows that Y is $\gamma(\frac{1}{2})$-variate.

Alternative method :

Here $Y = \frac{1}{2}X^2$ where X is a standard normal variate. The distribution function $\Phi(x)$ of X is given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt,\ -\infty < x < \infty.$$

Let $F(y)$ be the distribution function of Y.

Then $F(y) = P(Y \leqslant y)$ for all $y \in R$.

If $y < 0$, then $(Y \leqslant y) \Leftrightarrow (\frac{1}{2}X^2 \leqslant y)$, where $(\frac{1}{2}X^2 \leqslant y)$ is an impossible event. So $P(Y \leqslant y) = 0$, if $y < 0$, *i.e.*, $F(y) = 0$, if $y < 0$.

Now let $y \geqslant 0$. In this case

$P(Y \leqslant y) = P(\frac{1}{2}X^2 \leqslant \frac{1}{2}x^2) = P(X^2 \leqslant x^2) = P(-x \leqslant X \leqslant x)$, where $x > 0$.

$$= P(-\sqrt{2y} \leqslant X \leqslant \sqrt{2y}),\ (\because \text{ here } x = \sqrt{2y})$$

$$= P(-\sqrt{2y} < X \leqslant \sqrt{2y})$$

$$[X \text{ being continuous, } P(X = -\sqrt{2y}) = 0]$$

$$= \Phi(\sqrt{2y}) - \Phi(-\sqrt{2y}).$$

Thus we get $F(y) = 0$, if $y < 0$

$$= \Phi(\sqrt{2y}) - \Phi(-\sqrt{2y}),\ \text{if } y \geqslant 0.$$

Then if $f(y)$ be the density function of Y, we get

$$f(y) = F'(y) = 0 \quad \text{if } y < 0$$

and if $y > 0$,

$$f(y) = F'(y) = \frac{1}{\sqrt{2y}} \Phi'(\sqrt{2y}) + \frac{1}{\sqrt{2y}} \Phi'(-\sqrt{2y})$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2y}}(e^{-y} + e^{-y}),\ \left[\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right]$$

$$= \frac{1}{\sqrt{\pi}} e^{-y} y^{-\frac{1}{2}} = \frac{e^{-y} y^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})}$$

$f(0)$ can be defined ...... and we take $f(0) = 0$

Thus it is proved that the probability density function of Y is given by

$$f(y) = \frac{e^{-y} y^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})} \text{ if } 0 < y < \infty$$

$$= 0, \text{elsewhere.}$$

The above density function of Y shows that Y is a $\gamma(\frac{1}{2})$ variate.

Ex. 28. *In the equation $x^2 + 2x - q = 0$, q is a random variable uniformly distributed over the interval $(0, 2)$. Find the distribution function of the larger root.* [C. H. (Math.)'83, '88, '92]

Here q has uniform distribution over $(0, 2)$ So the density function $f_1(q)$ of q is [q within the paranthesis is regarded as a real variable] given by

$$f_1(q) = \tfrac{1}{2} \text{ if } 0 < q < 2$$

$$= 0, \text{ elsewhere.}$$

Now the larger root of $x^2 + 2x - q = 0$ is $-1 + \sqrt{1+q}$.

Let $Y = -1 + \sqrt{1+q}$. In real variable we get $y = -1 + \sqrt{1+q}$, where the real variable corresponding to q is also denoted by q.

Now $\dfrac{dy}{dq} = \dfrac{1}{2\sqrt{1+q}} > 0$ for all $q \in (0, 2)$.

Then the density function $f_Y(y)$ of Y is given by

$$f_Y(y) = \left|\frac{dq}{dy}\right| f_1(q).$$

So we get $f_Y(y) = 2\sqrt{1+q} \cdot \frac{1}{2}$ if $0 < q < 2$

or, $f_Y(y) = y + 1$ if $0 < y < \sqrt{3} - 1$

$$= 0, \text{ elsewhere.}$$

Let $F_Y(y)$ be the distribution function of Y.

Then $F_Y(y) = \displaystyle\int_{-\infty}^{y} f_Y(t)\, dt$, for all real values of y.

Now if $-\infty < y \leqslant 0$, then

$$F_Y(y) = \int_{-\infty}^{y} 0.dt = 0.$$

If $0 < y < \sqrt{3} - 1$, then

$$F_Y(y) = \int_{-\infty}^{0} f_Y(t)\,dt + \int_{0}^{y} f_Y(t)\,dt$$

$$= 0 + \int_0^y f_Y(t)\,dt$$

$$= \int_0^y (t+1)\,dt$$

$$= \frac{y^2}{2} + y.$$

Finally, if $y \geqslant \sqrt{3} - 1$, then

$$F_Y(y) = \int_{-\infty}^{0} f_Y(t)\,dt + \int_0^{\sqrt{3}-1} f_Y(t)\,dt + \int_{\sqrt{3}-1}^{y} f_Y(t)\,dt$$

$$= 0 + \int_0^{\sqrt{3}-1} (t+1)\,dt + 0$$

$$= 1.$$

So, the required distribution function $F_Y(y)$ is given by

$$F_Y(y) = 0, \quad \text{if } -\infty < y \leqslant 0$$

$$= \frac{y^2}{2} + y, \quad \text{if } 0 < y < \sqrt{3} - 1$$

$$= 1, \quad \text{if } y \geqslant \sqrt{3} - 1.$$

Ex. 29. *A point is chosen at random on a semi-circle having centre at the origin and radius unity and projected on the diameter. Prove that the distance of the point of projection from the centre has the probability density* $\dfrac{1}{\pi\sqrt{1-x^2}}$ *for* $-1 < x < 1$ *and zero elsewhere.*

[C.H. (Math.) '87]

We consider a semi-circle of unit radius. If O be the centre and P be the point of projection on the diameter AP of the point Q lying on the semi-circle, then the position of Q is determined by the angle Y, shown in the figure. The point Q being chosen at random, the density function of Y is given by

$$f_Y(y) = \frac{1}{\pi}, \; 0 < y < \pi.$$

Now $OP = X = \cos Y$.

In real variable $x = \cos y$, $0 < y < \pi$, $-1 < x < 1$.

$$\therefore \; \frac{dx}{dy} = -\sin y < 0 \quad \text{when} \quad 0 < y < \pi.$$

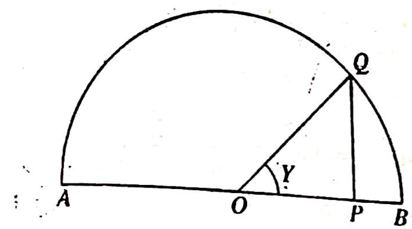


Fig. 5.14.6

Then $f_Y(y) = f_X(x) \left| \dfrac{dx}{dy} \right|$,

where $f_X(x)$ is the density function of X.

$$\therefore \; \frac{1}{\pi} = f_X(x)\sqrt{1-x^2}$$

i.e., the density function of X is given by

$$f_x(x) = \begin{cases} \dfrac{1}{\pi\sqrt{1-x^2}}, & -1 < x < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Ex. 30. *Find the distribution of the square of a Poisson μ-variate.*

Let X be Poisson μ-variate.

Then $P(X = i) = \dfrac{e^{-\mu}\mu^i}{i!}$, $i = 0, 1, 2, 3, \ldots\ldots$

Let $Y = X^2$. Then Y is also a discrete random variable having the spectrum,

$$\{0, 1^2, 2^2, \ldots\ldots\}, \text{ i.e., } \{i^2 : i = 0, 1, 2, 3, \ldots\ldots\},$$

$\therefore$ the probability distribution of Y is given by

$$P(Y = i^2) = P(X^2 = i^2) = P(X = i) = \frac{e^{-\mu}\mu^i}{i!}.$$

Hence the distribution of Y is given by the spectrum

$$y_i = i^3 \quad (i=0,\ 1,\ 2\ldots\ldots)$$

with $f_i = P(Y=i^3) = \dfrac{e^{-\mu}\mu^i}{i\,!}$.

Ex. 31. *If X is uniformly distributed in the interval $(-1, 1)$, then find the distribution of $|X|$.*

Let $Y=|X|$.

In real variable $y=|x|=\begin{cases} x, & x>0 \\ -x, & x<0 \\ 0, & x=0. \end{cases}$

$\therefore\ \dfrac{dy}{dx}=\begin{cases} 1, & x>0 \\ -1, & x<0, \end{cases}$

and $\dfrac{dy}{dx}$ does not exist for $x=0$. Also $\dfrac{dy}{dx}$ changes sign in $(-1, 1)$, where it exists.

Case I. Let $x>0$. We have

$$\begin{aligned} P(y<Y\leq y+dy) &= P(|x|<|X|\leq|x+dx|) \\ &= P(x<|X|\leq x+dx),\ x \text{ being positive} \\ &= P(x<X\leq x+dx) + P(-(x+dx)\leq X<-x) \\ &= 2P(x<X\leq x+dx) \end{aligned}$$

due to symmetry of the uniform distribution.

$\therefore\ f_Y(y)\,dy=2f_X(x)\,dx$, where $f_X(x)$ and $f_Y(y)$ are respectively the density functions of X and Y.

Now $f_X(x)=\frac{1}{2}$, $-1<x<1$.

$\therefore\ f_Y(y)=2f_X(x)\dfrac{dx}{dy}=\dfrac{dx}{dy}=1$ since $x>0$.

$\therefore\ f_Y(y)=1$, when x is positive and $0<y<1$.

Case II. Let $x<0$, then we have

$$\begin{aligned} P(y<Y\leq y+dy) &= P(|x|<|X|\leq|x+dx|) \\ &= P(-x<|X|\leq-(x+dx)),\ x \text{ being negative} \\ &= P\{(x+dx)\leq X<x\}+P\{-x<X\leq-(x+dx)\} \\ &= 2P(-x<X\leq-(x+dx)),\ \text{due to symmetry.} \end{aligned}$$

$\therefore\ f_Y(y)\,dy=-2f_X(x)\,dx.$

$\therefore\ f_Y(y)=-2f_X(x)\dfrac{dx}{dy}=-2\cdot\frac{1}{2}(-1)$, since $x<0$.

$=1$

Thus $f_Y(y)=1$ when $x<0$ and $0<y<1$.

Thus we see that in either case $f_Y(y)=\begin{cases} 1, & 0<y<1 \\ 0, & \text{elsewhere} \end{cases}$

i.e., $Y=|X|$ is uniformly distributed in $(0, 1)$.

Ex. 32. *The random variable X has the probability density function $f_X(x)=\frac{1}{2}e^{-|x|}$, $-\infty<x<\infty$. Find the probability density function of $Y=X^2$.*

Here $Y=X^2$. In real variable $y=x^2$, $-\infty<x<\infty$.

$\therefore\ \dfrac{dy}{dx}=2x$ which changes sign in $-\infty<x<\infty$. Then proceeding as in Ex. 31,

$$f_Y(y)\,dy=2f_X(x)\,dx, \text{ if } x>0,$$

where $f_Y(y)$ is the probability density function of Y.

Hence, if $x>0$, $f_Y(y)=\dfrac{2}{2x}f_X(x)=\dfrac{1}{2\sqrt{y}}e^{-\sqrt{y}}$, $0<y<\infty$.

It can be shown similarly that if $x<0$, then also we have

$$f_Y(y)=\frac{1}{2\sqrt{y}}\,e^{-\sqrt{y}},\ 0<y<\infty.$$

Thns we get, in either case,

$$f_Y(y)=\begin{cases} \dfrac{1}{2\sqrt{y}}\,e^{-\sqrt{y}}, & \text{if } y>0 \\ 0, & \text{elsewhere.} \end{cases}$$

Ex. 33. *The random variable X is uniformly distributed in $(0, 1)$. Find the distribution of $Y=-2\log_e X$.*

If $f_X(x)$ be the probability density function of X,

$$f_X(x)=\begin{cases} 1, & 0<x<1 \\ 0, & \text{elsewhere.} \end{cases}$$

Now, $Y=-2\log_e X$. In real variable $y=-2\log_e x$.

$\therefore\ \dfrac{dy}{dx}=-\dfrac{2}{x}<0$ for every $x\in(0, 1)$.

$\therefore$ the distribution of Y is given by the density function

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \tfrac{1}{2}e^{-\frac{1}{2}y},\ 0 < y < \infty.$$

Ex. 34. *The random variable X has the probability density function $f_X(x)$ given by*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \textit{elsewhere.} \end{cases}$$

Prove that $P(X > a+b \mid X > a) = P(X > b),\ a \geq 0,\ b \geq 0.$

Let $F(x)$ be the distribution function of X.

Then for $x \geq 0$, $F(x) = \int_0^x \lambda e^{-\lambda x}\,dx = 1 - e^{-\lambda x}$;

and $F(x) = 0$ for $x < 0$.

Now $P(X > a) = 1 - P(X \leq a) = 1 - F(a) = e^{-\lambda a}$ $a \geq 0$.

Similarly $P(X > b) = e^{-\lambda b}$, $P(X > a+b) = e^{-(a+b)\lambda}$,

where $a \geq 0,\ b \geq 0.$

$\therefore$ $P(X > a)\ P(X > b) = P(X > a+b).$

$\therefore$ $P(X > a+b \mid > X > a) = \dfrac{P(X > a+b)}{P(X > a)} = P(X > b).$

Ex. 35. *A random variable X has the following discrete distribution :*

x_i :	-3	-2	-1	0	1	2	3	4
f_i :	$2k^2$	k	$2k$	$3k$	$2k$	k	$7k^2$	k^2

Find k and also distribution of Y where

(i) $Y = X^2$, *(ii)* $Y = |X-1| + |X+1|$.

As in Ex. 12, $k = \frac{1}{10}$, so that the distribution of X is given by

x_i :	-3	-2	-1	0	1	2	3	4
f_i :	·02	·1	·2	·3	·2	·1	·07	·01.

(i) The distribution of Y is given by

$y_i = 0, 1, 4, 9, 16$ with

$f_1 = P(Y=0) = P(X=0 = {\cdot}3,$

$f_2 = P(Y=1) = P(X=1) + P(X=-1) = {\cdot}4$

$f_3 = P(Y=4) = P(X=2) + P(X=-2) = {\cdot}2$

$f_4 = P(Y=9) = P(X=3) + P(X=-3) = {\cdot}09$

$f_5 = P(Y=16) = P(X=4) = {\cdot}01.$

(ii) The distribution Y in this case, is given by

$y_i = 2, 4, 6, 8$ with

$f_1 = P(Y=2) = P(X=-1) + P(X=0) + P(X=1) = {\cdot}7$

$f_2 = P(Y=4) = P(X=-2) + P(X=2) = {\cdot}2$

$f_3 = P(Y=6) = P(X=-3) + P(X=3) = {\cdot}09$

$f_4 = P)Y=8) = P(X=4) = {\cdot}01.$

Ex. 36. *Let X be a random variable whose density function is given by* $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \textit{elsewhere.} \end{cases}$

Find the distribution of $Y = X^{\frac{1}{\beta}},\ \beta \neq 0.$

$Y = X^{\frac{1}{\beta}}$. In real variable $y = x^{\frac{1}{\beta}},\ \beta \neq 0,\ x > 0.$

Then $x = y^{\beta},\ \beta \neq 0,\ y > 0$ for every $x > 0$.

$\therefore$ $\dfrac{dx}{dy} = \beta y^{\beta-1} \gtrless 0$ according as $\beta \gtrless 0$, for every $y > 0$.

If $f_Y(y)$ be the density function of Y,

$$f_Y(y) = f(x)\left|\frac{dx}{dy}\right| = |\beta|\ \lambda y^{\beta-1}\ e^{-\lambda y^{\beta}},\ y > 0,$$

which gives the required distribution of Y.

Ex. 37. *Let X be a continuous random variable having density function $f(x)$. Prove that the density function of the random variable $Y = X^2$ is given by*

$$g(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left\{ f(\sqrt{y}) + f(-\sqrt{y}) \right\}, & y > 0 \\ 0, & \textit{elsewhere.} \end{cases}$$

Hence find the distribution of Y, when X is uniformly distributed in $(-1, 2)$.

Let $F(x)$ and $G(y)$ be the distribution functions of the random variables X and Y respectively.

If $y < 0$, the event $(Y \leq y) = (X^2 \leq y) = O$, the impossible event and so $G(y) = P(Y \leq y) = 0$ if $y < 0$.

MP-13

Also $G(0)=P(Y \leqslant 0)=P(X^2 \leqslant 0)=P(X=0)=0$, since X is a continuous variate.

Let, $y > 0$, then $G(y)=P(Y \leqslant y)=P(X^2 \leqslant y)$
$=P(-\sqrt{y} < X \leqslant \sqrt{y}) = F(\sqrt{y})-F(-\sqrt{y})$

$$\therefore \quad G'(y)=\frac{1}{2\sqrt{y}}\left[F'(\sqrt{y})+F'(-\sqrt{y})\right]$$

$$=\frac{1}{2\sqrt{y}}\left[f(\sqrt{y})+f(-\sqrt{y})\right],\ y > 0.$$

Also $G'(y)=0$ if $y < 0$ and $LG'(0)=0$, $RG'(0)=\infty$.

So $G'(0)$ does not exist. Since $g(0)$ can be defined arbitrarily, we take $g(0)=0$.

Hence the density function of $Y=X^2$ is given by

$$g(y)=\begin{cases} \frac{1}{2\sqrt{y}}\{f(\sqrt{y})+f(-\sqrt{y})\}, & y > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

If X be uniformly distributed in $(-1,2)$,

$$f(x)=\begin{cases} \frac{1}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Then $f(\sqrt{y})=\begin{cases} \frac{1}{3}, & 0 \leq \sqrt{y} < 2 \\ 0, & \text{elsewhere,} \end{cases}$

and $f(-\sqrt{y})=\begin{cases} \frac{1}{3}, & -1 < -\sqrt{y} \leqslant 0 \\ 0, & \text{elsewhere.} \end{cases}$

Now $0 \leqslant \sqrt{y} < 2$ implies $0 \leq y < 4$, and

$-1 < -\sqrt{y} \leqslant 0$ implies $0 \leq y < 1$.

Hence $f(\sqrt{y})=\begin{cases} \frac{1}{3}, & 0 \leqslant y < 4 \\ 0, & y \geqslant 4 \end{cases}$

and $f(-\sqrt{y})=\begin{cases} \frac{1}{3}, & 0 \leq y < 1 \\ 0, & y \geqslant 1 \end{cases}$

Hence $g(y)=\begin{cases} \frac{1}{2\sqrt{y}}\left(\frac{1}{3}+\frac{1}{3}\right), & 0 < y \leq 1 \\ \frac{1}{2\sqrt{y}}\left(\frac{1}{3}+0\right), & 1 < y < 4. \end{cases}$

i.e., $g(y)=\begin{cases} \frac{1}{3\sqrt{y}}, & 0 < y \leqslant 1 \\ \frac{1}{6\sqrt{y}}, & 1 < y \leq 4. \end{cases}$

Ex. 38. *If X has gamma distribution with parameter l, find the distribution of $Y=\frac{1}{2}X^2$.*

X being gamma distributed with parametere l, the density function $f_X(x)$ is given by

$$f_X(x)=\begin{cases} \frac{e^{-x}x^{l-1}}{\Gamma(l)}, & 0 < x < \infty,\ l > 0. \\ 0, & \text{elsewhere.} \end{cases}$$

Now $Y=\frac{1}{2}X^2$. In real variable $y=\frac{1}{2}x^2$.

$$\therefore \quad \frac{dy}{dx}=x > 0 \text{ for every } x > 0.$$

$\therefore$ The distribution of Y is given by the density function,

$$f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|=\frac{e^{-\sqrt{2y}}(2y)^{\frac{l}{2}-1}}{\Gamma(l)},\ y > 0.$$

Ex. 39. *If the random variable X is $\beta_2(m, n)$ distributed, prove that the random variable $Y=\frac{1}{X}$ is $\beta_2(n, m)$ distributed.*

X being $\beta_2(m, n)$ distributed, its density function $f_X(x)$ is given by,

$$f_X(x)=\begin{cases} \frac{x^{m-1}}{B(m, n)(1+x)^{m+n}}, & 0 < x < \infty,\ m > 0,\ n > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Now $Y=\frac{1}{X}$. In real variable, $y=\frac{1}{x}$.

$$\therefore \quad \frac{dy}{dx}=-\frac{1}{x^2} < 0 \text{ for every } x \in (0, \infty).$$

Hence the density function $f_Y(y)$ of Y is given by

$$f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|=\frac{x^{m+1}}{B(m, n)(1+x)^{m+n}}$$

$$=\frac{y^{n-1}}{B(m, n)(1+y)^{m+n}},\quad 0 < y < \infty.$$

Hence Y is a $\beta_2(n, m)$ variate.

Ex. 40. *If a real number X be chosen at random from $[1, 50]$ then find the probability that $X+\frac{125}{X} > 40$.*

X being a real number chosen at random from the interval $[1, 50]$ the corresponding random variable X is uniformly distributed with density function $f(x)$ given by

$$f(x)=\begin{cases} \frac{1}{49}, & 1 \leqslant x \leqslant 50 \\ 0, & \text{elsewhere.} \end{cases}$$

Now $X+\frac{125}{X}>40 \Leftrightarrow (X-20)^2>(5\sqrt{11})^2$

$\Leftrightarrow |X-20|>5\sqrt{11} \Leftrightarrow X>20+5\sqrt{11}$ or, $X<20-5\sqrt{11}$

Thus the required probability

$=P\left(X+\frac{125}{X}>40\right)=1-P\left(|X-20|\leq 5\sqrt{11}\right)$

$=1-P\left(20-5\sqrt{11}\leq X\leq 20+5\sqrt{11}\right)$

$=1-\int_{20-5\sqrt{11}}^{20+5\sqrt{11}} f(x)\,dx=1-\frac{1}{49}\int_{20-5\sqrt{11}}^{20+5\sqrt{11}} dx=1-\frac{10\sqrt{11}}{49}=1-0.68=0.32$

Ex. 41. *The weight of students in a college is normally distributed with $m=40$ kg and $\sigma=5$ kg. Find the percentage of the students that have weight (a) greater than 40 kg. (b) greater than 50 kg. (c) between 38 kg and 52 kg.*

$$\left[\text{Given that } \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{2} e^{-\frac{t^2}{2}}\,dt=\cdot 9772,\ \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\cdot 4} e^{-\frac{t^2}{2}}\,dt=\cdot 6554\right.$$

$$\left.\text{and } \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{2\cdot 4} e^{-\frac{t^2}{2}}\,dt=\cdot 9918.\right]$$

Let X be the random variable corresponding to the weight of a student in kg. Then X is normal (40, 5).

(a) $P(X>40)=P\left(\frac{X-40}{5}>0\right)=P(Z>0)=\cdot 5,$

since Z is the standard normal variate. Hence the required percentage is 50%.

(b) $P(X>50)=P\left(\frac{X-40}{5}>2\right)=P(Z>2)$

$$=1-P(Z\leq 2)=1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{2} e^{-\frac{t^2}{2}}\,dt=\cdot 0228$$

Hence the required percentage is 2·3% (approximately).

(c) $P(38<X<52)=P\left(\frac{38-40}{5}<\frac{X-40}{5}<\frac{52-40}{5}\right)$

$=P(-\cdot 4<Z<2\cdot 4)$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{2\cdot 4} e^{-\frac{t^2}{2}}\,dt-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\cdot 4} e^{-\frac{t^2}{2}}\,dt$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{2\cdot 4} e^{-\frac{t^2}{2}}\,dt+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\cdot 4} e^{-\frac{t^2}{2}}\,dt-1=\cdot 6472.$$

Hence the required percentage is 64·7%.

Ex. 42. *The length of bolts produced by a machine is normally distributed with parmeters $m=4$ and $\sigma=0\cdot 5$. A bolt is defective if its length does not lie in the interval (3·8, 4·3). Find the percentage of defective bolts produced by the machine.*

$$\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\cdot 6} e^{-\frac{t^2}{2}}\,dt=\cdot 7257,\ \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\cdot 4} e^{-\frac{t^2}{2}}\,dt=\cdot 6554.\right]$$

Let X be the random variable corresponding to the length of a bolt produced by the machine. Then X is normal (4, 0·5).

The required probability $=1-P(3\cdot 8<X<4\cdot 3)$

$=1-P\left(\frac{3\cdot 8-4}{0\cdot 5}<\frac{X-4}{0\cdot 5}<\frac{4\cdot 3-4}{0\cdot 5}\right)$

$=1-P(-\cdot 4<Z<\cdot 6)$

where Z is the standard normal variate.

$=1-[\Phi(\cdot 6)-\Phi(-\cdot 4)]$

$=2-[\Phi(\cdot 6)+\Phi(\cdot 4)]$

$$=2-\left[\int_{-\infty}^{\cdot 6} e^{-\frac{t^2}{2}}\,dt+\int_{-\infty}^{\cdot 4} e^{-\frac{t^2}{2}}\,dt\right]$$

$=2-(\cdot 7257+\cdot 6554)=\cdot 62$ (approximately)

Hence 62% of the bolts produced are defective.

Ex. 43. The random variable X is normally distributed with parameters m and σ such that

$$P(9{\cdot}8 < X < 14{\cdot}6) = 0{\cdot}4514.$$

Find m and σ, given that $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\cdot 6} e^{-\frac{t^2}{2}}\,dt = 0{\cdot}7257.$

If Z be the standard normal variate,

$$P(-0{\cdot}6 < Z < 0{\cdot}6) = \Phi(0{\cdot}6) - \Phi(-0{\cdot}6)$$

$$= 2\Phi(0{\cdot}6) - 1 = \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\cdot 6} e^{-\frac{t^2}{2}}\,dt - 1$$

$$= 2\times 0{\cdot}7257 - 1 = 0{\cdot}4514$$

$$P(9{\cdot}8 < X < 14{\cdot}6) = P\left(\frac{9{\cdot}8 - m}{\sigma} < Z < \frac{14{\cdot}6 - m}{\sigma}\right).$$

A solution for m, σ is given by

$$\therefore \quad \frac{9{\cdot}8 - m}{\sigma} = -0{\cdot}6 \text{ and } \frac{14{\cdot}6 - m}{\sigma} = 0{\cdot}6.$$

Solving $m = 12{\cdot}2$ and $\sigma = 4$.

Ex. 44. A straight line is drawn through a fixed point (α, β) $(\alpha > 0)$ making an angle X, which is chosen at random in the interval $(0, \pi)$, with the y-axis. Prove that the intercept on the y-axis, Y has a Cauchy distribution with parameters α, β.

The equation of any line which makes an angle X $(0 < X < \pi)$ with the y-axis and cuts off an intercept Y from that axis is

$$y = x \cot X + Y.$$

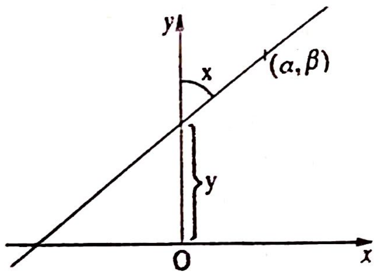


Fig. 5.14.7

If this line passes through (α, β), $Y = \beta - \alpha \cot X$, and the two random variables X, Y are connected by this relation. The density

function $f_X(x)$ of the random variable X, being uniformly distributed in $(0, \pi)$, is given by

$$f_X(x) = \begin{cases} \dfrac{1}{\pi}, & 0 < x < \pi \\ 0, & \text{elsewhere.} \end{cases}$$

Now in real variable, $y = \beta - \alpha \cot x$.

$$\therefore \quad \frac{dy}{dx} = \alpha \operatorname{cosec}^2 x > 0 \text{ for every } x \in (0, \pi).$$

Also as x varies from 0 to π, y varies from $-\infty$ to ∞.

The density function $f_Y(y)$ of Y is given by

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{\pi\,\alpha \operatorname{cosec}^2 x} = \frac{\alpha}{\pi[\alpha^2 + (y-\beta)^2]}, \quad -\infty < y < \infty.$$

Hence Y has Cauchy distribution with parameters α, β.

Ex. 45. A point X is chosen at random on a line segment AB whose middle point is O. Find the probability that AX, BX and AO form the sides of a triangle.

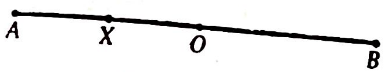


Fig. 5.14.8

Let $AB = 2a$ and X be any point taken at random on the segment AB. If $AX = Y$, then Y is uniformly distributed having density function $f(y)$ given by

$$f(y) = \begin{cases} \dfrac{1}{2a}, & 0 < y < 2a \\ 0, & \text{elsewhere.} \end{cases}$$

If O be the mid-point of AB, then the segments AX, BX and AO will form the sides of a triangle if

$$\left.\begin{array}{l} AX + BX > AO, \\ AX + AO > BX, \\ \text{and} \quad BX + AO > AX \end{array}\right\} \quad \begin{array}{l} \text{i.e., if } 2a > a,\ Y > \frac{a}{2},\ Y < \frac{3a}{2} \\ \text{i.e., if } \frac{a}{2} < Y < \frac{3a}{2}. \end{array}$$

$$\therefore \quad \text{the required probability} = \int_{\frac{a}{2}}^{\frac{3a}{2}} f(y)\,dy = \frac{1}{2}.$$

Ex. 46. *A point P is chosen at random on a line segment AB of length 2a. Find the probability that the area of the rectangle AB,PB will exceed $\frac{1}{2}a^2$.*

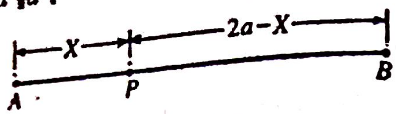


Fig. 5.14.9

Let $AP=X$. Then X is uniformly distributed with density function $f(x)$ given by $f(x)=\frac{1}{2a}$, if $0 < x < 2a$

$=0$, elsewhere.

The required probability $=P(X(2a-X) > \frac{1}{2}a^2)$

$$=P(X^2-2aX+\tfrac{1}{2}a^2 < 0)$$

$$=P\left\{\left(X-a+\frac{a}{\sqrt{2}}\right)\left(X-a-\frac{a}{\sqrt{2}}\right) < 0\right\}$$

$$=P\left(a-\frac{a}{\sqrt{2}} < X < a+\frac{a}{\sqrt{2}}\right)$$

$$=\int_{a-\frac{a}{\sqrt{2}}}^{a+\frac{a}{\sqrt{2}}} f(x)dx=\frac{1}{\sqrt{2}}.$$

Ex. 47. *X is a Poisson variate with parameter μ. Show that $P(X < n)=\frac{1}{n!}\int_\mu^\infty e^{-x}x^n\,dx$, where n is any positive integer.*

[C. H. (Math.) '94]

We have $P(X < n)=P(X=0)+P(X=1)+\cdots\cdots+P(X=n)$

$$=e^{-\mu}+e^{-\mu}\mu+\cdots\cdots+\frac{e^{-\mu}\mu^n}{n!}$$

Now $\displaystyle\int_\mu^\infty e^{-x}x^n\,dx.$

$$=\underset{B\to\infty}{Lt}\left[-x^n e^{-x}\right]_\mu^B+\underset{B\to\infty}{Lt}\int_\mu^B nx^{n-1}e^{-x}\,dx$$

$$=\mu^n e^{-\mu}+n\underset{B\to\infty}{Lt}\int_\mu^B e^{-x}x^{n-1}\,dx,$$

since $\underset{B\to\infty}{Lt}\frac{B^n}{e^B}=0$, n being a positive integer.

So we get

$$\frac{1}{n!}\int_\mu^\infty e^{-x}x^n\,dx=\frac{e^{-\mu}\mu^n}{n!}+\frac{1}{(n-1)!}\int_\mu^\infty e^{-x}x^{n-1}\,dx.$$

Similarly, we find that

$$\frac{1}{(n-1)!}\int_\mu^\infty e^{-x}x^{n-1}\,dx=\frac{e^{-\mu}\mu^{n-1}}{(n-1)!}+\frac{1}{(n-2)!}\int_\mu^\infty e^{-x}x^{n-2}\,dx.$$

$$\cdots\qquad\cdots\qquad\cdots\qquad\cdots$$

$$\cdots\qquad\cdots\qquad\cdots\qquad\cdots$$

$$\frac{1}{1!}\int_\mu^\infty e^{-x}x\,dx=\frac{e^{-\mu}\mu}{1!}+\int_\mu^\infty e^{-x}\,dx=\frac{e^{-\mu}\mu}{1!}+e^{-\mu}.$$

Adding we get

$$\frac{1}{n!}\int_\mu^\infty e^{-x}x^n\,dx=\frac{e^{-\mu}\mu^n}{n!}+\frac{e^{-\mu}\mu^{n-1}}{(n-1)!}+\cdots\cdots+\frac{e^{-\mu}\mu}{1!}+e^{-\mu}.$$

So it is proved that $P(X\leq n)=\frac{1}{n!}\int_\mu^\infty e^{-x}x^n\,dx.$

Ex. 48. *A point chosen at random in a given interval of length a divides it into two subintervals. Find the probability that the ratio of the length of the left side interval to that of the right subinterval is less than a constant K.*

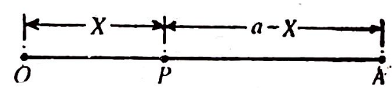


Fig. 5.14.10

Let X be the random variable denoting the length of the left side interval. The density function of the random variable X is

$$f(x)=\begin{cases}\frac{1}{a}, & 0 < x < a\\ 0, & \text{elsewhere.}\end{cases}$$

Let $Y=\frac{X}{a-X}$.

In terms of real variable $y=\frac{x}{a-x}$.

$\therefore \frac{dy}{dx}=\frac{a}{(a-x)^2}>0$ for every $x \in (0, a)$.

Also as x varies from 0 to a, y varies from 0 to $+\infty$.

$\therefore$ the density function $f_Y(y)$ of Y is given by

$$f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|=\frac{(a-x)^2}{a^2}=\frac{1}{(1+y)^2}, \quad 0<y<\infty.$$

$\therefore$ required probability$=P(Y<K)=\int_0^K \frac{dy}{(1+y)^2}=\frac{K}{1+K}$.

Ex. 49. *If X is normal (0, 1), then find the distribution of e^X.*

X being normal (0, 1), its density function $f_X(x)$ is given by

$$f_x(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad -\infty<x<\infty.$$

Let $Y=e^X$.

In terms of real variable, $y=e^x$.

$\therefore \frac{dy}{dx}=e^x>0$, for all $x \in (-\infty, \infty)$.

As x varies from $-\infty$ to ∞, y varies from 0 to ∞.

$\therefore$ the distribution of Y is given by the density function

$$f_Y(y)=f_X(x)\cdot\left|\frac{dx}{dy}\right|=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}e^{-x}$$

$$=\frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{(\log y)^2}{2}}}{y}, \quad 0<y<\infty.$$

Ex. 50. *Suppose that a projectile is fired at an angle X with the horizontal and with a velocity u. X has uniform distribution in $\left(\frac{\pi}{6}, \frac{\pi}{4}\right)$. Find the probability density function of the horizontal range R of the projectile.*

Let R be the random variable corresponding to the horizontal range of the particle with initial velocity u at an angle X with the horizon. The probability density function $f_X(x)$ of X is given by

$$f_X(x)=\frac{12}{\pi}, \text{ if } \frac{\pi}{6}<x<\frac{\pi}{4}$$
$$=0, \text{ elsewhere.}$$

Writing $R=\frac{u^2}{g}\sin 2X$, in real variable we get

$$r=\frac{u^2}{g}\sin 2x.$$

Then $\frac{dr}{dx}=\frac{2u^2}{g}\cos 2x>0$ if $\frac{\pi}{6}<x<\frac{\pi}{4}$.

Now when x varies from $\frac{\pi}{6}$ to $\frac{\pi}{4}$, r varies from $\frac{\sqrt{3}u^2}{2g}$ to $\frac{u^2}{g}$.

So, if $f_R(r)$ be the probability density function of R, then

$$f_R(r)=\frac{\frac{12}{\pi}g}{2u^2\cos 2x} \text{ if } \frac{\pi}{6}<x<\frac{\pi}{4}$$

$$=\frac{6g}{\pi u^2\cos 2x}=\frac{6g}{\pi u^2\sqrt{1-\left(\frac{gr}{u^2}\right)^2}}$$

Hence, $f_R(r)=\frac{6g}{\pi\sqrt{u^4-g^2r^2}}, \quad \frac{\sqrt{3}u^2}{2g}<r<\frac{u^2}{g}$.

Ex. 51. *The probability density function of the random variable X is given by*

$$f(x)=\begin{cases} 2xe^{-x^2}, & x>0 \\ 0, & \text{elsewhere.}\end{cases}$$

Find the distribution of X^2.

We put $Y=X^2$.

In terms of real variable, $y=x^2$.

$\therefore \frac{dy}{dx}=2x>0$ for all $x>0$.

Hence if $f_Y(y)$ be the density function of Y,

$$f_Y(y)=2xe^{-x^2}\frac{1}{|2x|}=e^{-x^2} \text{ when } x>0.$$

So $\quad f_Y(y)=e^{-y}, \; y>0,$

which determines the distribution of Y.

Ex. 52. *If X is a $\gamma(l)$ variate, then find the distribution of $\sqrt{X}$.*

X being a $\gamma(l)$ variate, its density function is given by

$$f_X(x)=\begin{cases}\dfrac{e^{-x}x^{l-1}}{\Gamma(l)}, & 0<x<\infty,\ l>0\\ 0 & , \text{ elsewhere.}\end{cases}$$

Now $Y=\sqrt{X}$. In real variable $y=\sqrt{x}$

$$\therefore\ \frac{dy}{dx}=\frac{1}{2\sqrt{x}}>0 \text{ for every } x>0.$$

As x varies from 0 to ∞, y varies from 0 to ∞.

Hence the distribution of Y is given by the density function

$$f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|=\frac{e^{-x}x^{l-1}}{\Gamma(l)}2\sqrt{x}$$

$$=\frac{2e^{-y^2}y^{2l-1}}{\Gamma(l)},\ 0<y<\infty.$$

Ex. 53. *The probability that a bacterium which is alive till time t measured from its instant of birth will die between time t and $t+\Delta t$ is $\lambda\Delta t+0(\Delta t)$ where λ is a constant and $\dfrac{0(\Delta t)}{\Delta t}\to 0$ as $\Delta t\to 0$.*

Assuming that the probability of death at the instant of birth is zero, find the probability that the bacterium is dead before its lifespan becomes t. [C. H. (Math.) '95]

Let $A(t)$ and B denote respectively the events 'bacterium is dead before its lifespan becomes t', 'becterium is dead between t and $t+\Delta t$'.

Let $P[A(t)]=f(t)$.

Then $P[A(t+\Delta t)]=f(t+\Delta t)$

It is given that the conditional probability

$$P[B\mid\overline{A(t)}]=\lambda\Delta t+0(\Delta t).$$

We are to find the value of $P[A(t)]$, *i.e.*, of $f(t)$.

We observe that $A(t+\Delta t)$ happens if and only if $A(t)$ happens or $\overline{A(t)}\,B$ happens so that we can write $A(t+\Delta t)=A(t)+\overline{A(t)}\ B$ where $A(t)$, $\overline{A(t)}\,B$ are mutually exclusive events.

So $P[A(t+\Delta t)]=P[A(t)]+P[\overline{A(t)}\ B]$

or, $f(t+\Delta t)=f(t)+P[\overline{A}(t)]P[B\mid\overline{A(t)}]$

or, $f(t+\Delta t)-f(t)=[1-f(t)]\cdot[\lambda\Delta t+0(\Delta t)]$

or, $$\frac{f(t+\Delta t)-f(t)}{\Delta t}=\lambda[1-f(t)]+[1-f(t)]\frac{0(\Delta t)}{\Delta t}\quad\cdots\cdots(5.14.2)$$

Taking limits of both sides of (5.14.2) as $\Delta t\to 0$ we get

$$f'(t)=\lambda[1-f(t)]$$

or, $$f'(t)+\lambda f(t)=\lambda\quad\cdots\quad\cdots\quad(5.14.3)$$

Solving the linear differential equation (5.14.3) we get

$$f(t)e^{\lambda t}=e^{\lambda t}+c$$

or, $f(t)=ce^{-\lambda t}+1$ where c is a constant.

Now it is given that $f(0)=0$. $\therefore\ c=-1$.

So $f(t)=1-e^{-\lambda t}$.

Hence the required probability is $1-e^{-\lambda t}$.

Ex. 54. *Examine whether the following functions are distribution functions:*

(*i*) $F_1(x)=\dfrac{2}{\pi}\tan^{-1}x,\ -\infty<x<\infty$

(*ii*) $F_2(x)=0$ *if* $x\le 1$
$\qquad=1-\dfrac{1}{2x}$ *if* $x>1$

(*iii*) $F_3(x)=\dfrac{1}{2h}\displaystyle\int_{x-h}^{x+h}F(t)\,dt,\ -\infty<x<\infty,$

where $h(\neq 0)$ is a given constant and $F(x)$ is a distribution function.

If can be shown that a real valued function $F(x)$ of a real variable x defined in $(-\infty,\infty)$ is the distribution function of a random variable with respect to a suitable probability space (S,Δ,P) if

(*a*) $F(x)$ is monotonically increasing in $(-\infty,\infty)$,

(*b*) $F(\infty)=1$,

(*c*) $F(-\infty)=0$,

(*d*) $F(x)$ is continuous to the right of every point a.

For (*i*), we see that $F_1(-\infty)=\underset{x\to-\infty}{Lt}\dfrac{2}{\pi}\tan^{-1}x=-1\neq 0$.

So $F_1(x)$ is not a distribution function.

For (*ii*), we see that $F_2(1+0)=\frac{1}{2}$ and $F_2(1)=0$, so that $F_2(1+0) \neq F_2(1)$ and hence, $F_2(x)$ is not a distribution function.

For (*iii*), $h>0$ or $h<0$. Let $h>0$. $F(t)$, being a distribution function, $F(t)$ is monotonically increasing in $[x-h, x+h]$.

Then $\frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt$ is well defined for all $x \in (-\infty, \infty)$.

Now $\frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt=\frac{1}{2h}\int_{-h}^{h} F(x+y)\,dy$, where $t=x+y$.

So, $F_3(x)=\frac{1}{2h}\int_{-h}^{h} F(x+y)\,dy$, for all $x \in (-\infty, \infty)$.

Let $x_2 > x_1$. Then $x_2+y > x_1+y$ and $F(x_2+y) \geqslant F(x_1+y)$ for all $y \in [-h, h]$,

Further $\int_{-h}^{h} F(x_2+y)\,dy$ and $\int_{-h}^{h} F(x_1+y)\,dy$ both exist and so

$$\int_{-h}^{h} F(x_2+y)\,dy \geqslant \int_{-h}^{h} F(x_1+y)\,dy$$

$$\therefore \frac{1}{2h}\int_{-h}^{h} F(x_2+y)\,dy \geqslant \frac{1}{2h}\int_{-h}^{h} F(x_1+y)\,dy.$$

$(\because$ here $h>0)$

Hence, $F_3(x_2) \geqslant F_3(x_1)$, whenever $x_2 > x_1$.

So, $F_3(x)$ is monotonically increasing in $(-\infty, \infty)$.

Again, from the expression for $F_3(x)$ given by

$$F_3(x)=\frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt,$$

we find that $F_3(x)$ is continuous in $(-\infty, \infty)$ and hence $F_3(x)$ is continuous to the right of every point a.

Now $F_3(x)$ being monotonically increasing in $(-\infty, \infty)$, we have $F(x-h) \leqslant F(t) \leqslant F(x+h)$ of all $t \in [x-h, x+h]$.

$$\therefore \int_{x-h}^{x+h} F(x-h)\,dt \leqslant \int_{x-h}^{x+h} F(t)\,dt \leqslant \int_{x-h}^{x+h} F(x+h)\,dt$$

$$\text{or, } 2hF(x-h) \leqslant \int_{x-h}^{x+h} F(t)\,dt \leqslant 2hF(x+h)$$

$$\therefore F(x-h) \leqslant \frac{1}{2h}\int_{x-h}^{x+h} F(t)\,dt \leqslant F(x+h). \quad [\because \text{ here } h>0]$$

Thus we get $F(x-h) \leqslant F_3(x) \leqslant F(x+h)$, for all $x \in (-\infty, \infty)$.

Now F being a distribution function we have

$$\underset{x\to\infty}{Lt} F(x-h)=\underset{x\to\infty}{Lt} F(x+h)=1,$$

and $\underset{x\to-\infty}{Lt} F(x-h)=\underset{x\to-\infty}{Lt} F(x+h)=0$, and hence we conclude that $\underset{x\to\infty}{Lt} F_3(x)=1$ and $\underset{x\to-\infty}{Lt} F_3(x)=0$. So it is proved that $F_3(x)$ satisfies the conditions (*a*), (*b*), (*c*), (*d*) and so $F_3(x)$ is a distribution function when $h>0$. It can be proved similarly that $F_3(x)$ is a distribution function when $h<0$.

Examples V

1. Consider the random experiment of to sing of 2 coins. Find the distribution of the number of heads.

2. Can the following be probability mass functions or density functions ?

(*i*) $f(x)=\begin{cases} \cdot 2, & x=-3 \\ \cdot 3, & x=-2 \\ \cdot 4, & x=0 \\ \cdot 1, & x=1 \\ 0, & \text{elsewhere.} \end{cases}$ (*ii*) $f(x)=\begin{cases} \frac{x}{2}, & 0<x\leqslant 1 \\ \frac{1}{2}, & 1<x\leqslant 2 \\ \frac{3-x}{2}, & 2<x\leqslant 3 \\ 0, & \text{elsewhere.} \end{cases}$

3. Find the value of the constant k, so that the function $f(x)$ defined below may be probability density function :

(*i*) $f(x)=\begin{cases} x, & 0<x\leqslant 1 \\ k-x, & 1<x\leqslant 2 \\ 0, & \text{elsewhere.} \end{cases}$ (*ii*) $f(x)=\begin{cases} kx^3, & 0<x<1 \\ 0, & \text{elsewhere.} \end{cases}$

4. X has the probability density function

$$f(x)=\begin{cases} \frac{1}{4}, & -2<x<2 \\ 0, & \text{elsewhere.} \end{cases}$$

Obtain (*i*) $P(X<1)$, (*ii*) $P(|X|>1)$, (*iii*) $P(2X+3>5)$.

5. Let X be a random variable having probability density function given by

$$f(x)=\begin{cases} cx, & 0 \leq x \leq 1 \\ c, & 1 \leq x \leq 2 \\ -cx+3c, & 2 \leq x \leq 3 \\ 0, & \text{elsewhere.} \end{cases}$$

Find c and the distribution function.

6. A random variable X has the following distribution :

$x_i = i$:	-2	-1	0	1	2	3
$f_i = P(X=i)$:	$\cdot 15$	k	$\cdot 25$	$2k$	$\cdot 35$	k

Find the value of k and obtain the distribution function $F(x)$.

7. Consider the random experiment of drawing a ball from an urn containing red and white balls. Let p be the probability of drawing a red ball. Define a random variable corresponding to the experiment and find its distribution.

8. Verify that the function $f(x)=\begin{cases} \frac{1}{\theta}\, e^{-\frac{x}{\theta}}, & 0 < x < \infty \\ 0, & \text{elsewhere} \end{cases}$

is a possible probability density function. Find $P(2 < X < 6)$.

9. Find the probability density function of the distribution, whose distribution function $F(x)$ is given by

(i) $$F(x)=\begin{cases} \frac{x^2}{2}, & 0 < x \leq 1 \\ -\frac{1}{8}+\frac{1}{4}\left(3x-\frac{x^2}{2}\right), & 1 < x \leq 3 \\ 1, & x > 3, \\ 0, & \text{elsewhere.} \end{cases}$$

(ii) $$F(x)=\begin{cases} 0, & x < 0 \\ \frac{x}{8}, & 0 \leq x < 1 \\ \frac{1}{8}, & 1 \leq x \leq 2 \\ \frac{x}{8}, & 2 \leq x < 4 \\ 1, & x \geq 4. \end{cases}$$

10. A fair coin is tossed three times. The three random variables X, Y, Z are defined as follows :

X denotes the number of heads, Y denotes the number of occurrences of at least two consecutive heads, Z denotes the initial

run of heads, *i.e.*, the number of heads preceding the first tail. Find the probability distributions of X, Y, Z.

[*Hints* :

Outcomes	X	Y	Z
(H, H, T)	2	1	2
(H, H, H)	3	1	3
(T, T, T)	0	0	0
(H, T, H)	2	0	1
(T, T, H)	1	0	0
(T, H, T)	1	0	0
(H, T, T)	1	0	1
(T, H, H)	2	1	0.

From above we find that spectra of X, Y, Z are respectively $\{0, 1, 2, 3\}$, $\{0, 1\}$, $\{0, 1, 2, 3\}$.

$P(X=0)=P\{(T, T, T)\}=\frac{1}{8}$,

$P(X=1)=P\{(T, T, H)\}+P\{(T, H, T)\}+P\{(H, T, T)\}=\frac{3}{8}$,

$P(X=2)=\frac{3}{8}$, $P(X=3)=\frac{1}{8}$, $P(Y=0)=\frac{5}{8}$, $P(Y=1)=\frac{3}{8}$,

$P(Z=0)=\frac{1}{2}$, $P(Z=1)=\frac{1}{4}$, $P(Z=2)=\frac{1}{8}$, $P(Z=3)=\frac{1}{8}$.]

11. Prove that $f(x)=\frac{1}{2}\, e^{-|x|}$, $-\infty < x < \infty$ is a possible probability density function. Find the corresponding distribution function.

[*Hints* : Since $|x|$ is an even function and $|x|=x$ for every $x > 0$,

$$\int_{-\infty}^{\infty} f(x)\, dx=\tfrac{1}{2}\int_{-\infty}^{\infty} e^{-|x|}\, dx=\int_{0}^{\infty} e^{-x}\, dx$$

$$=1, \quad f(x) \geq 0 \text{ for every } x.$$

Hence $f(x)$ is a possible probability density function. Let $F(x)$ be the corresponding distribution function.

In $-\infty < x < 0$, $F(x)=\int_{-\infty}^{x} \frac{1}{2}e^{t}\, dt=\frac{e^{x}}{2}$.

In $0 \leq x < \infty$, $F(x)=\int_{-\infty}^{x} f(t)\, dt=\int_{-\infty}^{0} f(t)\, dt+\int_{0}^{x} f(t)\, dt$

$$=\tfrac{1}{2}+\frac{-e^{-x}+1}{2}=1-\tfrac{1}{2}e^{-x} \,]$$

12. A discrete random variable X has the following probability mass function :

$x_i = i$:	-3	-2	-1	0	1	2
$f_i = P(X=i)$:	k	$2k$	$2k^2$	$3k^2$	k^2	$6k^2+8k$.

(i) Determine the value of k. (ii) Find the distribution function $F(x)$. (iii) Evaluate $P(X<-1)$.

[*Hints* : (i) $\sum_i f_i = 1$, $k > 0$, gives $k = \frac{1}{12}$.

(ii) The distribution of X then becomes

$x_i = i$:	-3	-2	-1	0	1	2
$f_i = P(X=i)$:	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{72}$	$\frac{1}{48}$	$\frac{1}{144}$	$\frac{17}{24}$.

If $F(x)$ be the distribution function,

in $-\infty < x < -3$, $F(x) = 0$,

in $-3 \leqslant x < -2$, $F(x) = P(X=-3) = \frac{1}{12}$,

in $-2 \leqslant x < -1$, $F(x) = P(X=-3) + P(X=-2) = \frac{1}{4}$,

in $-1 \leqslant x < 0$, $F(x) = P(X=-3) + P(X=-2) + P(X=-1) = \frac{19}{72}$

in $0 \leqslant x < 1$, $F(x) = \frac{41}{144}$,

in $1 \leqslant x < 2$, $F(x) = \frac{7}{24}$,

in $2 \leqslant x < \infty$, $F(x) = 1$,

(iii) $P(X<-1) = \frac{1}{4}$.]

13. The probability density function of a random variable X is symmetric about the origin. Prove that X and $-X$ have the same distribution.

[*Hints* : Let $f(x)$ be the probability density function of X.

$f(x)$ being symmetric, $f(-x) = f(x)$ for every x.

Now $P(-\infty < -X \leqslant x) = P(X \geqslant -x)$

$$= \int_{-x}^{\infty} f(t)\,dt = \int_{-x}^{\infty} f(t)\,dt = \underset{B\to\infty}{Lt} \int_{-x}^{B} f(t)\,dt$$

$$= \underset{B\to\infty}{Lt} \int_{-B}^{x} f(-y)\,dy, \text{ where } t = -y$$

$$= \int_{-\infty}^{x} f(-y)\,dy = \int_{-\infty}^{x} f(y)\,dy = P(X \leqslant x).$$

Hence X and $-X$ have the same distribution.]

Ex. V

14. For what values of θ and c, is the function f_i given by

$$f_i = \begin{cases} \dfrac{c\theta^i}{i}, & i \text{ is a positive integer} \\ 0, & \text{elsewhere} \end{cases}$$

a possible probability mass function ?

[*Hints* : $\sum_i f_i = 1$, which implies $c\left(\theta + \frac{\theta^2}{2} + \frac{\theta^3}{3} + \cdots\cdots\right) = 1$

i.e., $-c \log(1-\theta) = 1$ provided $-1 \leqslant \theta < 1$.

$\therefore\ c = -\dfrac{1}{\log(1-\theta)}$, $-1 \leqslant \theta < 1$, give the required values of c and θ.]

15. A bomber flies directly above a railway track. Assume that if a large bomb falls within 30 ft of the track, the track will be sufficiently damaged and that the traffic will be disrupted. Let X denote the perpendicular distance from the track of any point where a bomb falls. The probability density function of X is given by

$$f(x) = \begin{cases} \dfrac{150-x}{10^4}, & 0 \leqslant x \leqslant 200 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the probability that the bomb will disrupt the traffic.

(b) If the plane can carry three bombs and uses all the three bombs, what is the probability that traffic will be disrupted ?

[*Hints* : (a) Required probability $= \displaystyle\int_0^{30} \frac{150-x}{10^4} = \cdot 405$.

(b) Assuming that the three bombs attack the track independently, the probability that all the three bombs fail to hit the target $= (\cdot 595)^3$.

Hence the required probability $= 1 - (\cdot 595)^3 = \cdot 789$.]

16. Show that the distribution of a random variable X, whose distribution function $F(x)$ is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \leqslant x < 1 \\ \frac{3}{4}, & 1 \leqslant x < 2 \\ 1, & 2 \leqslant x \end{cases}$$

is a mixed distribution.

17. Suppose that the number of defective screws produced by a sophisticated machine per day has a Poisson distribution with parameter 2. What are the probabilities that out of the total produce of a day, there are (*i*) no defective screw, (*ii*) exactly 2 defective screws, (*iii*) at least one defective screw, (*iv*) less than 3 defective screws ?

18. The km run (in thousand km) which car-owners get with a certain brand of type is a random variable with probability density function

$$f(x)=\begin{cases}\frac{1}{40}e^{-\frac{x}{40}}, & x>0\\ 0, & \text{elsewhere.}\end{cases}$$

Find the probability that one of these tyres will last (i) at most 15,000 kms., (*ii*) between 40,000 and 50,000 kms, (*iii*) at least 10,000 kms.

[*Hints* : If $F(x)$ be the distribution function

$$F(x)=\begin{cases}0, & -\infty<x<0\\ \int_0^x f(x)\,dx=1-e^{-\frac{x}{40}}, & 0\leqslant x<\infty.\end{cases}$$

(*i*) Required probability $=P(X\leqslant 15)=F(15)=1-e^{-\frac{3}{8}}$

(*ii*) Required probability$=(40\leqslant X\leqslant 50)$

$$=F(50)-F(40)=e^{-1}-e^{-\frac{5}{4}}.$$

(*iii*) Required probability$=P(X\geqslant 10)=1-P(X<10)$

$$=1-F(10)=1-(1-e^{-\frac{1}{4}})=e^{-\frac{1}{4}}.]$$

19. The number of emergency admission each day to a hospital is found to have a Poisson distribution with parameter 2. (*a*) Evaluate the probability that on a particular day there will be no emergency admissions. (*b*) At the beginning of one day the hospital has five beds available for emergency. Calculate the probability that this will be an insufficient number for the day. [*C.H.* (*Math.*) '92]

[*Hints* : The probability for *i* admissions on any day

$$=P(X=i)=\frac{e^{-2}2^i}{i\,!},\ i \text{ is a non-negative integer.}$$

Ex. V

(*a*) Required probability $=P(X=0)=e^{-2}$.

(*b*) Required probability$=P(X>5)=1-P(X<5)$

$$=1-e^{-2}\left(1+2+\frac{2^2}{2!}+\frac{2^3}{3!}+\frac{2^4}{4!}+\frac{2^5}{5!}\right)\simeq \cdot 0166.]$$

20. An office switch-board receives phone calls at the rate of 2 in every 5 minutes on the average. What is the probability of getting exactly 4 calls in 15 minutes ?

[*Hints* : Taking 5 minutes as unit of time we see that there are 3 units of time in the interval of 15 minutes and so we have $\lambda=2$, $t=3$. If X be the random variable denoting the number of phone calls in the given interval, then X has Poisson distribution with parameter $\lambda t=6$.

The required probability $=P(X=4)=\dfrac{e^{-6}6^4}{4\,!}\simeq \cdot 1339.]$

21. A machine starts producing products of which 5% are defective. The producer draws 10 products every hour for inspection. If the sample does not contain any defective item, the machine is not stopped. Find the probability that the machine will not be stopped.

[*Hints* : If X be the random variable denoting the number of defective items out of 10 products drawn, then X has the binomial distribution :

$$f_i=P(X=i)=\binom{10}{i}\left(\tfrac{1}{20}\right)^i\left(\tfrac{19}{20}\right)^{10-i},\ i=0, 1, 2, \ldots\ldots 10.$$

Required probability$=\left(\frac{19}{20}\right)^{10}$.]

22. The probability that a family should have at least one boy and at least one girl is at least ·90. At least how many children should the family have ?

[*Hints* : n = number of children. If X be the random variable denoting the number of boys in the family, then X has the binomial distribution, where $f_i=P(X=i)=\binom{n}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{n-i}$, $i=0, 1, 2,\ldots\ldots n$. Now the probability of the event "the family has at least one boy and at least one girl"

$$=P(1\leqslant X\leqslant n-1)=1-\{P(X=0)+P(X=n)\}=1-2\cdot\frac{1}{2^n}.$$

By the given condition $1-\dfrac{1}{2^{n-1}}\geqslant \cdot 90$, *i.e.*, least value of n is 5.]

23. A number is randomly chosen from the interval [0, 1]. What is the probability that (a) its first decimal digit will be 2, (b) its second decimal digit will be 4, (c) the first decimal digit of its square root is 5 ?

[*Hints*: The distribution of the random variable denoting the number chosen from the interval [0, 1] is given by the probability density function

$$f(x)=\begin{cases}1, & 0 \leq x \leq 1\\ 0, & \text{elsewhere.}\end{cases}$$

(a) Required probability $=P(\cdot 2 \leq X < \cdot 3)=\int_{\cdot 2}^{\cdot 3} dx=\cdot 1.$

(b) Required probability $=\sum_{k=0}^{9} P(0{\cdot}k4 \leq X < 0{\cdot}k5)$

$$=\sum_{k=0}^{9}\int_{0\cdot k4}^{0\cdot k5} dx=\sum_{k=0}^{9}\cdot 01=\cdot 1.$$

(c) Required probability $=P(\cdot 5 \leq \sqrt{X} \leq \cdot 6)$
$=P(\cdot 25 \leq X < \cdot 36)=\cdot 11.$]

24. A certain airline company having observed 5% of the persons seeking reservations on a flight do not show up for the flight, sells 100 seats on a plane that has 95 seats. What is the probability that there will be a seat available for every person who shows up for the flight ?

[*Hints*: Let X be the random variable denoting the number of persons showing up for the flight.

Required probability $=P(X \leq 95)=1-P(X>95)$

$$=1-\sum_{k=96}^{100}\binom{100}{k}(\cdot 95)^k(\cdot 05)^{100-k}.\ ]$$

25. The random variable X is normal (50, 20). Find $P(\,|\,X-50\,|\leq 20)$, given that $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1} e^{-\frac{x^2}{2}}dx=\cdot 8413.$

[*Hints*: $P(\,|\,X-50\,|\leq 20)=P\left(-1 \leq \frac{X-50}{20} \leq 1\right)$

$=\Phi(1)-\Phi(-1)=2\Phi(1)-1=2\times\cdot 8413-1=\cdot 6826.$]

26. The I.Q. of students of a class is normally distributed with parameters $m=100$ and $\sigma=10$. If the total number of students in the class is 700, then find the number of students who have I.Q. $>$ 115. Given that $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1\cdot 5} e^{-\frac{t^2}{2}}dt=\cdot 9332.$

[*Hints*: If X be the random variable denoting the I.Q. of students, then X is normal (100, 10).

Now $P(X>115)=P\left(\frac{X-100}{10}>1{\cdot}5\right)=P(Z>1{\cdot}5)$

$$=1-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{1\cdot 5} e^{-\frac{t^2}{2}}dt=\cdot 0668.$$

$\therefore$ required number of students $=700\times\cdot 0668=47$ nearly.]

27. The weight of students in a college is normally distributed with $m=40$ kg and $\sigma=5$ kg. In a random sample of 5 students, what is the probability that (*i*) all will have weights greater than 50 kg, (*ii*) 3 will have weights greater than 50 kg ?

[*Hints*: Let p be the probability that a student selected at random has weight over 50 kg. This probability p remains the same for every student chosen at random (Illustrative Ex. 41(*b*)).

(*a*) Required probability $=(\cdot 0228)^5$,

(*b*) Required probability $=\binom{5}{3}(\cdot 0228)^3(\cdot 9772)^2$

by binomial law with $n=5$, $p=\cdot 0228$.]

28. The parameter b in the quadratic equation $x^2-2bx+1=0$ is distributed normally ($\cdot 6$, 1). Find the probability that the roots of the quadratic are real.

$\left[\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt=\cdot 1554\right.$ and $\cdot 4452$ for $x=\cdot 4$ and $x=1{\cdot}6$ respectively]. [C.H. (Math.) '71]

[*Hints*: Since the parameter b is normally distributed ($\cdot 6$, 1), its density function is given by

$$f(b)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(b-\cdot 6)^2}{2}},\ -\infty<b<\infty,$$

Now the roots of the equation $x^2-2bx+1=0$ will be real if $b^2>1$, *i.e.*, if $b>1$ or $b\leqslant -1$ and the two events $(b\leqslant -1)$ and $(b>1)$ are mutually exclusive.

$$\therefore\ P(b^2>1)=P(b>1)+P(b\leqslant -1)=\int_1^{\infty} f(b)\,db+\int_{-\infty}^{-1} f(b)\,db$$

$$=\frac{1}{\sqrt{2\pi}}\left(\int_1^{\infty} e^{-\frac{(b-\cdot 6)^2}{2}}\,db+\int_{-\infty}^{-1} e^{-\frac{(b-\cdot 6)^2}{2}}\,db\right)$$

$$=\frac{1}{\sqrt{2\pi}}\left(\underset{B_1\to\infty}{Lt}\int_1^{B_1} e^{-\frac{(b-\cdot 6)^2}{2}}\,db+\underset{B_2\to-\infty}{Lt}\int_{B_2}^{-1} e^{-\frac{(b-\cdot 6)^2}{2}}\,db\right)$$

$$=\frac{1}{\sqrt{2\pi}}\left(\underset{B_1\to\infty}{Lt}\int_{\cdot 4}^{B_1-\cdot 6} e^{-\frac{t^2}{2}}\,dt+\underset{B_2\to-\infty}{Lt}\int_{1\cdot 6}^{-B_2+\cdot 6} e^{-\frac{z^2}{2}}\,dz\right)$$

where $t=b-\cdot 6$ and $z=-(b-\cdot 6)$

$$=\frac{1}{\sqrt{2\pi}}\left(\int_{\cdot 4}^{\infty} e^{-\frac{t^2}{2}}\,dt+\int_{1\cdot 6}^{\infty} e^{-\frac{z^2}{2}}\,dz\right)$$

$$=\frac{1}{\sqrt{2\pi}}\left(\int_0^{\infty} e^{-\frac{t^2}{2}}\,dt-\int_0^{\cdot 4} e^{-\frac{t^2}{2}}\,dt+\int_0^{\infty} e^{-\frac{t^2}{2}}\,dt\right.$$

$$\left.-\int_0^{1\cdot 6} e^{-\frac{t^2}{2}}\,dt\right)$$

$$=\frac{1}{\sqrt{2\pi}}\left(2\int_0^{\infty} e^{-\frac{t^2}{2}}\,dt-\int_0^{\cdot 4} e^{-\frac{t^2}{2}}\,dt-\int_0^{1\cdot 6} e^{-\frac{t^2}{2}}\,dt\right)$$

$$=\frac{1}{\sqrt{2\pi}}\left(2\times\sqrt{\frac{\pi}{2}}-\cdot 1554-\cdot 4452\right)=\cdot 3994.\ ]$$

29. The random variable X is normal (m, σ). Find the distribution of $Y=aX+b$, where a, b are constants and $a\neq 0$.

30. A random variable X has the following distribution :

x_i :	-2	-1	0	1	2
f_i :	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{11}{30}$

Find the distribution of $Y=X^2$.

31. X is uniformly distributed in (2, 6). Find the probability that $Y<10$, where $Y=X^2-6$.

32. If X is a binomial (n, p) variate, then find the distribution of $Y=aX+b$, where a, b are constants and $a\neq 0$.

[*Hints* : The spectrum of Y is given by

$$y_i=ai+b,\quad i=0, 1, 2, \ldots\ldots n.$$

$$P(X=i)=P(aX+b=ai+b)=P(Y=y_i)$$

i.e., $f_{yi}=P(Y=y_i)=P(X=i)=\binom{n}{i}p^i(1-p)^{n-i}$.]

33. If X be uniformly distributed in (0, 1), then find the distribution of $Y=X^{\frac{1}{\beta}}$, $\beta\neq 0$.

34. If X is $\beta_1(l, m)$ variate, then find the distribution of the random variable $Y=\frac{X}{1-X}$.

Answers

1. $X=i$, $i=0, 1, 2$ with $P(X=0)=\frac{1}{4}$, $P(X=1)=\frac{1}{2}$, $P(X=2)=\frac{1}{4}$.
2. (*i*) *p.m.f.* (*ii*) *p.d.f.*
3. (*i*) $k=2$. (*ii*) $k=4$.
4. (*i*) $\frac{3}{4}$.

   (*ii*) $(|X|>1)=P(X>1)+P(X<-1)$

   $$=\int_1^2 f(x)dx+\int_{-2}^{-1} f(x)dx=\frac{1}{2}$$

   (*iii*) $P(2X+3>5)=P(X>1)=\frac{1}{4}$.

5. $C=\frac{1}{2}$, $F(x)=\begin{cases}0, & x<0\\ \frac{x^2}{4}, & 0\leq x\leq 1\\ \frac{2x-1}{4}, & 1<x\leqslant 2\\ \frac{6x-x^2-5}{4}, & 2<x\leqslant 3\\ 1, & x>3.\end{cases}$

6. $k=\frac{1}{16}$, $F(x)=\begin{cases}0, & x<-2\\ \frac{3}{80}, & -2\leq x<-1\\ \frac{17}{80}, & -1\leq x<0\\ \frac{37}{80}, & 0\leq x<1\\ \frac{47}{80}, & 1\leq x<2\\ \frac{75}{80}, & 2\leq x<3\\ 1, & x\geqslant 3.\end{cases}$

7. X is defined as follows : $X=1$, if red ball is drawn, $X=0$, if white ball is drawn.

$X=0, 1$ with $P(X=0)=1-p,\ P(X=1)=p.$

$$F(x)=\begin{cases} 0 & , x<0 \\ 1-p, & 0 \leqslant x < 1 \\ 1 & , x \geqslant 1. \end{cases}$$

8. $e^{-\frac{2}{\theta}}-e^{-\frac{6}{\theta}}.$

9. (i) $f(x)=\begin{cases} x, 0<x \leqslant 1 \\ \frac{1}{2}(3-x),\ 1<x \leqslant 3 \\ 0, \text{ elsewhere.} \end{cases}$

(ii) $f(x)=\begin{cases} 0, x<0 \\ \frac{1}{2}, 0 \leqslant x<1 \\ 0, 1 \leqslant x<2 \\ \frac{1}{4}, 2 \leqslant x<4 \\ 0, x \geqslant 4. \end{cases}$

17. (i) $\dfrac{e^{-2}.2^0}{0!}=\cdot 135$, (ii) $\dfrac{e^{-2}.2^2}{2!}=\cdot 270$,

(iii) $1-P(X=0)=\cdot 865$,

(iv) $P(X=0)+P(X=1)+P(X=2)=\cdot 540.$

29. Y is normal $(am+b, \mid a \mid \sigma)$.

30.

y_i :	0	1	4
$P(Y=y_i)$:	$\frac{1}{5}$	$\frac{7}{30}$	$\frac{17}{30}$.

31. $f_Y(y)=\dfrac{1}{8\sqrt{y+6}}$, $-2<y<30$, $P(Y<10)=\frac{1}{2}$.

33. $f_Y(y)=\begin{cases} \mid \beta \mid y^{\beta-1} & , 0<y<1 \\ 0 & , \text{elsewhere.} \end{cases}$

34. $f_Y(y)=\dfrac{y^{l-1}}{B(l, m)(1+y)^{l+m}}$, $0<y<\infty$ i.e., Y is a $\beta_2(l, m)$ variate.

DISTRIBUTIONS OF MORE THAN ONE DIMENSION

6.1. Multidimensional Random Variables.

In the previous chapter we have restricted our discussion to the probability distribution of a single random variable defined on the event space. We also know that many random variables can be defined on the same event space. In many situations it will be convenient to consider the joint probability distribution (explained in the next section) of several random variables instead of considering the probability distribution of a single random variable. We give an example. Let E be the random experiment of throwing a die and then tossing a coin. The event space S of E is given by

$$S=\{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H), (1, T), \ldots, (6, T)\},$$

which contains 12 distinct outcomes. A random variable X is defined on S as follows :

$$X(i, H)=i \text{ for } i=1, 2, \ldots\ldots, 6,$$
$$X(i, T)=i \text{ for } i=1, 2, \ldots\ldots, 6.$$

Also let another random variable Y be defined on S as follows :

$$Y(i, H)=0,\ Y(i, T)=-1, \text{ for } i=1, 2, \ldots\ldots, 6.$$

Then the above two random variables can be described as follows :

X denotes the number on the die and

$Y=0$ if 'head' appears,

$Y=-1$ if 'tail' appears.

Here we see that the event 'head appears' cannot be described solely by X and the event 'number on the die is i' cannot be described solely by Y. Here it will be convenient to consider the two-dimensional random variable (X, Y) which can be defined as follows :

$$(X, Y)(\omega)=(X(\omega), Y(\omega)), \text{ for all } \omega \in S.$$

Then $(X, Y)(2, H)=(X(2, H), Y(2, H))=(2, 0)$,

$(X, Y)(6, T)=(X(6, T), Y(6, T))=(6, -1)$ etc.

Now $(X, Y)(2, H)=(2, 0)$ can be described as $(X=2, Y=0)$ which can be read as 'joint occurrence of the events $(X=2)$ and $(Y=0)$'. In general if S be an event space and if X, Y are random variables defined on S, then the *two-dimensional random variable*

7. X is defined as follows: $X=1$, if red ball is drawn, $X=0$, if white ball is drawn.

$X=0, 1$ with $P(X=0)=1-p$, $P(X=1)=p$.

$$F(x)=\begin{cases} 0 & , x<0 \\ 1-p, & 0 \leq x < 1 \\ 1 & , x \geq 1. \end{cases}$$

8. $e^{-\frac{2}{\theta}} - e^{-\frac{6}{\theta}}$.

9. (i) $f(x)=\begin{cases} x, 0 < x \leq 1 \\ \frac{1}{4}(3-x), 1 < x \leq 3 \\ 0, \text{ elsewhere.} \end{cases}$

(ii) $f(x)=\begin{cases} 0, x<0 \\ \frac{1}{2}, 0 \leq x < 1 \\ 0, 1 \leq x < 2 \\ \frac{1}{4}, 2 \leq x < 4 \\ 0, x \geq 4. \end{cases}$

17. (i) $\dfrac{e^{-2}.2^0}{0!}=\cdot 135$, (ii) $\dfrac{e^{-2}.2^2}{2!}=\cdot 270$,

(iii) $1-P(X=0)=\cdot 865$,

(iv) $P(X=0)+P(X=1)+P(X=2)=\cdot 540$.

29. Y is normal $(am+b, |a| \sigma)$.

30.

y_i:	0	1	4
$P(Y=y_i$:	$\frac{1}{5}$	$\frac{7}{30}$	$\frac{17}{30}$.

31. $f_Y(y)=\dfrac{1}{8\sqrt{y+6}}$, $-2 < y < 30$, $P(Y < 10)=\frac{1}{2}$.

33. $f_Y(y)=\begin{cases} |\beta| y^{\beta-1}, & 0<y<1 \\ 0 & , \text{ elsewhere.} \end{cases}$

34. $f_Y(y)=\dfrac{y^{l-1}}{B(l, m)(1+y)^{l+m}}$, $0 < y < \infty$ i.e., Y is a $\beta_2(l, m)$ variate.

CHAPTER VI

DISTRIBUTIONS OF MORE THAN ONE DIMENSION

6.1. Multidimensional Random Variables.

In the previous chapter we have restricted our discussion to the probability distribution of a single random variable defined on an event space. We also know that many random variables can be defined on the same event space. In many situations it will be convenient to consider the joint probability distribution (explained in the next section) of several random variables instead of considering the probability distribution of a single random variable. We give an example. Let E be the random experiment of throwing a die and then tossing a coin. The event space S of E is given by

$$S=\{(1, H), (2, H), (3, H), (4, H), (5, H), (6, H), (1, T), ..., (6, T)\},$$

which contains 12 distinct outcomes. A random variable X is defined on S as follows:

$$X(i, H)=i \text{ for } i=1, 2, ..., 6,$$
$$X(i, T)=i \text{ for } i=1, 2, ..., 6.$$

Also let another random variable Y be defined on S as follows:

$$Y(i, H)=0, \quad Y(i, T)=-1, \quad \text{for } i=1, 2, ..., 6.$$

Then the above two random variables can be described as follows:

X denotes the number on the die and

$Y=0$ if 'head' appears,

$Y=-1$ if 'tail' appears.

Here we see that the event 'head appears' cannot be described solely by X and the event 'number on the die is i' cannot be described solely by Y. Here it will be convenient to consider the two-dimensional random variable (X, Y) which can be defined as follows:

$$(X, Y)(\omega)=(X(\omega), Y(\omega)), \text{ for all } \omega \in S.$$

Then $(X, Y)(2, H)=(X(2, H), Y(2, H))=(2, 0)$,

$(X, Y)(6, T)=(X(6, T), Y(6, T))=(6, -1)$ etc.

Now $(X, Y)(2, H)=(2, 0)$ can be described as $(X=2, Y=0)$ which can be read as 'joint occurrence of the events $(X=2)$ and $(Y=0)$'. In general if S be an event space and if X, Y are random variables defined on S, then the *two-dimensional random variable*

(X, Y) can be defined as a mapping $(X, Y) : S \rightarrow R^2$, where for every $\omega \in S$, $(X, Y)(\omega) = (X(\omega), Y(\omega)) \in R^2$. The range of this mapping is called the *spectrum of the two-dimensional random variable* (X, Y). For two real numbers a, b, $(X=a, Y=b)$ represents the event $\{\omega : \omega \in S$, where $(X, Y)(\omega) = (a, b)\}$ and it represents 'simultaneous occurrence of the events $(X=a)$ and $(Y=b)$'.

For the real numbers a, b, c, d where $a<b$, $c<d$, the event

$$\{\omega : \omega \in S \text{ where } a < X(\omega) \leqslant b,\ c < Y(\omega) \leqslant d\}$$

is expressed as $(a < X \leqslant b, c < Y \leqslant d)$ and it represents 'simultaneous occurrence of the events $a < X \leqslant b$ and $c < Y \leqslant d$'.

If n random variables $X_1, X_2, X_3, .., X_n$ be defined on the same event space S, then the *n-dimensional random variable*

$(X_1, X_2, ..., X_n)$ can be defined as a mapping

$$(X_1, X_2, ..., X_n) : S \rightarrow R^n,$$

where for every $\omega \in S$,

$$(X_1, X_2, .., X_n)(\omega) = (X_1(\omega), X_2(\omega), ..., X_n(\omega)) \in R^n.$$

Then $(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$ represents simultaneous occurrence of the events

$$(X_1 = x_1), (X_2 = x_2), ..., (X_n = x_n)$$

and $(a_1 < X_1 \leqslant b_1, a_2 < X_2 \leqslant b_2, ..., a_n < X_n \leqslant b_n)$ represents the simultaneous occurrence of the events

$$(a_1 < X_1 \leqslant b_1), (a_2 < X_2 \leqslant b_2), ..., (a_n < X_n \leqslant b_n).$$

6.2. Distribution function in more than one dimension.

Let X and Y be two random variables defined on the same event space S and let P be a given probabilty function defined on a given class of subsets (of S) forming the class Δ of events. The joint distribution function of two random variables X, Y, or the *distribution function of the two dimensional random varirble* (X, Y), is a function $F : R \times R \rightarrow R$ defined by

$$F(x, y) = P(-\infty < X \leqslant x, -\infty < Y \leqslant y), \text{ for all } x, y \in R, \quad (6.2.1)$$

where, as before, the event $(-\infty < X \leqslant x, -\infty < Y \leqslant y)$ denotes the simultaneous occurrence of the events $(-\infty < X \leqslant x)$ and $(-\infty < Y \leqslant y)$. As usual we shall say that $F(x, y)$ is a distribution function instead of speaking F is a distribution function.

Similarly the *distribution function of the n-dimensional random variable* $(X_1, X_2, ..., X_n)$ or the joint distribution function of $X_1, X_2, .., X_n$ is defined by

$$F(x_1, x_2, ..., x_n) = P(-\infty < X_1 \leqslant x_1, -\infty < X_2 \leqslant x_2, ..., -\infty < X_n \leqslant x_n) \quad (6.2.2)$$

for all $x_1, x_2, ..., x_n \in R$, and is called an *n-dimensional distribution function*.

To illustrate the concept let us consider the random experiment of throwing two dice. Let X and Y denote the numbers shown by the first and second die respectively. The outcomes of this experiment are shown by dots in Fig 6.2.1. The corresponding event space contains 36 simple events which are assumed to be equally likely.

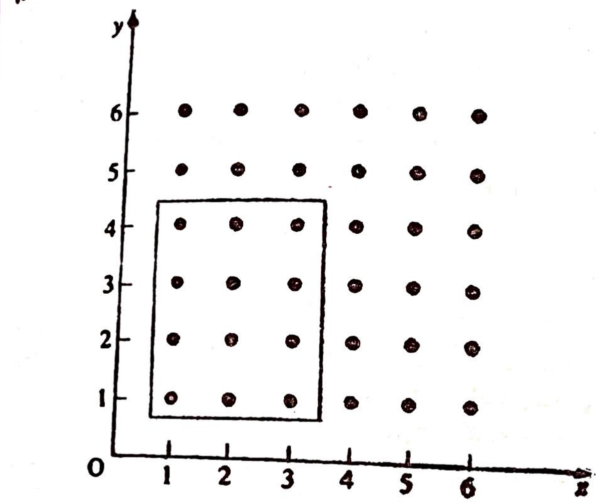


Fig 6.2.1. Outcomes of the random experiment of throwing a pair of dice.

If $F(x, y)$ be the distribution function of the two dimensional random variable (X, Y), we can find $F(x, y)$ at each point (x, y).

For example,

$$F(3, 4) = P(-\infty < X \leqslant 3, -\infty < Y \leqslant 4)$$
$$= \frac{12}{36} = \frac{1}{3},$$

since the event $(-\infty < X \leqslant 3, -\infty < Y \leqslant 4)$ contains 12 outcomes shown within the rectangular block in Fig 6.2.1.

Properties of Two Dimensional Distribution Function.

I. $F(x, y)$ is monotonically non-decreasing in both the variables x and y.

Let y be kept fixed and $x_2 > x_1$.

Then, $(-\infty < X \leqslant x_2, -\infty < Y \leqslant y)$

$$=(-\infty < X \leqslant x_1, -\infty < Y \leqslant y)+(x_1 < X \leqslant x_2, -\infty < Y \leqslant y)$$

and since the last two events are mutually exclusive,

$$P(-\infty < X \leqslant x_2, -\infty < Y \leqslant y)$$
$$=P(-\infty < X \leqslant x_1, -\infty < Y \leqslant y)+P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y).$$

$\therefore\ F(x_2, y)=F(x_1, y)+P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y).$

Since $P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y) \geqslant 0$, it follows that

$$F(x_2, y) \geqslant F(x_1, y), \text{whenever } x_2 > x_1.$$

$\therefore\ F(x, y)$ is monotonically non-decreasing in x. Similarly it can be shown that $F(x, y_2) \geqslant F(x, y_1)$, whenever $y_2 > y_1$. So $F(x, y)$ is monotonically non-decreasing in y also.

II. $0 \leqslant F(x, y) \leqslant 1$ for all real values of x and y. (6.2.3)

We have for all real values of x and y,

$$0 \leqslant P(-\infty < X \leqslant x, -\infty < Y \leqslant y) \leqslant 1$$

or, $0 \leqslant F(x, y) \leqslant 1.$

III. $P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$

$$=F(x_2, y_2)-F(x_1, y_2)-F(x_2, y_1)+F(x_1, y_1), \tag{6.2.4}$$

where $x_2 > x_1$ and $y_2 > y_1$.

We have $(-\infty < X \leqslant x_2, -\infty < Y \leqslant y_2)$

$$=(-\infty < X \leqslant x_1, -\infty < Y \leqslant y_2)+(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_2)$$

and the last two events being mutually exclusive,

So, $P(-\infty < X \leqslant x_2, -\infty < Y \leqslant y_2)$

$$=P(-\infty < X \leqslant x_1, -\infty < Y \leqslant y_2)+P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_2)$$

or, $F(x_2, y_2)=F(x_1, y_2)+P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_2).$ (6.2.5)

Similarly, $F(x_2, y_1)=F(x_1, y_1)$
$$+P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_1). \tag{6.2.6}$$

Again $(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_2)$

$$=(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_1)+(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$$

and the last two events being mutually exclusive,

So, $P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_2)$
$$=P(x_1 < X \leqslant x_2, -\infty < Y \leqslant y_1)+P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$$

or, by (6.2.5) and (6.2.6),

$$F(x_2, y_2)-F(x_1, y_2)$$
$$=F(x_2, y_1)-F(x_1, y_1)+P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2).$$

$\therefore\ P(x_1 < X \leqslant x_2, y_1 < Y \leqslant y_2)$
$$=F(x_2, y_2)-F(x_1, y_2)-F(x_2, y_1)+F(x_1, y_1).$$

IV. $F(\infty, \infty)=1,\ F(-\infty, y)=0=F(x, -\infty).$ (6.2.7)

$F(\infty, \infty)=\underset{x\to\infty}{Lt}\ \underset{y\to\infty}{Lt}\ F(x, y)=\underset{y\to\infty}{Lt}\ \underset{x\to\infty}{Lt}\ F(x, y)$, where we note that the repeated limits exist, since $F(x, y)$ is monotonically non-decreasing in each of x and y and $F(x, y)$ is bounded in $R \times R$.

We consider the sequence of events $\{A_n\}$ and $\{B_m\}$ defined by $A_n=(-\infty < X \leqslant n)$ and $B_m=(-\infty < Y \leqslant m)$, where n and m are positive integers.

Since $\underset{x\to\infty}{Lt}\ \underset{y\to\infty}{Lt}\ F(x, y)$ exists, we have

$$\underset{x\to\infty}{Lt}\ \underset{y\to\infty}{Lt}\ F(x, y)=\underset{n\to\infty}{Lt}\ \underset{m\to\infty}{Lt}\ F(n, m)$$
$$=\underset{n\to\infty}{Lt}\ \underset{m\to\infty}{Lt}\ P(-\infty < X \leqslant n, -\infty < Y \leqslant m).$$

Now $(-\infty < X \leqslant n, -\infty < Y \leqslant m)=A_nB_m$ where A_nB_m represents the simultaneous occurrence of the events A_n and B_m. We keep n fixed.

Now $B_m \subseteq B_{m+1}$ for all m.

$\therefore\ A_nB_m \subseteq A_nB_{m+1}$ for all m.

Let $C_m=A_nB_m$.

Then $C_m \subseteq C_{m+1}$ for all m. So $\{C_m\}$ is a monotonically non-decreasing sequence of events,

$$\text{whence } \underset{m\to\infty}{Lt}\ C_m=\sum_{m=1}^{\infty} C_m.$$

$$\text{Then } P(\underset{m\to\infty}{Lt}\ C_m)=\underset{m\to\infty}{Lt}\ P(C_m)$$

$$\text{or, } P\left(\sum_{m=1}^{\infty} C_m\right)=\underset{m\to\infty}{Lt}\ P(A_n\,B_m) \tag{6.2.8}$$

or, $P\left(\sum_{m=1}^{\infty} C_m\right) = \underset{m\to\infty}{Lt}\ F(n, m)$ where n is fixed.

Now $\sum_{m=1}^{\infty} C_m = A_n\left(\sum_{m=1}^{\infty} B_m\right)$. (6.2.9)

We now show that $\sum_{m=1}^{\infty} B_m = S$, the corresponding event space.

$\omega \in \sum_{m=1}^{\infty} B_m$ implies $\omega \in B_m$ for some m and so $\omega \in S$, since each B_m is a subset of S.

$$\therefore \quad \sum_{m=1}^{\infty} B_m \subseteq S. \tag{6.2.10}$$

Again $\omega \in S$ implies that $Y(\omega)$ is a definite real number, say c. Now by Archimedean property, there exists a positive integer m_1 such that $1.m_1 > c$, *i.e.*, $Y(\omega) < m_1$ and consequently, $\omega \in (-\infty < Y \leq m_1) = B_{m_1}$.

$$\therefore \quad \omega \in \sum_{m=1}^{\infty} B_m.$$

$$\text{Hence } S \subseteq \sum_{m=1}^{\infty} B_m. \tag{6.2.11}$$

From (6.2.10) and (6.2.11), we get $\sum_{m=1}^{\infty} B_m = S$.

Then from (6.2.9), $\sum_{m=1}^{\infty} C_m = A_n S = A_n$.

$\therefore \quad P(A_n) = \underset{m\to\infty}{Lt}\ F(n, m)$, when n is fixed.

But n can be fixed arbitrarily.

$$\text{So } \underset{m\to\infty}{Lt}\ F(n, m) = P(A_n) \text{ for all } n. \tag{6.2.12}$$

Again $A_n \subseteq A_{n+1}$ for all n. So $\{A_n\}$ is a monotonically non-decreasing sequence of events, whence

$$\underset{n\to\infty}{Lt}\ A_n = \sum_{n=1}^{\infty} A_n.$$

Then $P(\underset{n\to\infty}{Lt}\ A_n) = \underset{n\to\infty}{Lt}\ P(A_n)$

$$\text{or, } P\left(\sum_{n=1}^{\infty} A_n\right) = \underset{n\to\infty}{Lt}\ P(A_n). \tag{6.2.13}$$

It can be proved, as before, that $\sum_{n=1}^{\infty} A_n = S$.

Then from (6.2.12), and (6.2.13), we get

$$\underset{n\to\infty}{Lt}\ \underset{m\to\infty}{Lt}\ F(n, m) = \underset{n\to\infty}{Lt}\ P(A_n) = P(S) = 1.$$

Hence $F(\infty, \infty) = 1$.

It can be similarly shown that

$$F(-\infty, y) = F(x, -\infty) = 0.$$

$$\text{V. } F(a+0, c) = F(a, c),\ F(a, c+0) = F(a, c). \tag{6.2.14}$$

We consider sequence of events $\{A_n\}$ and an event B, where $A_n = \left(a < X \leq a + \frac{1}{n}\right)$ and $B = (-\infty < Y \leq c)$, where n is a positive integer. We see that

$A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots\ldots$ and hence

$A_1B \supseteq A_2B \supseteq A_3B \supseteq \ldots\ldots$, *i.e.*, the sequence $\{C_n\}$, where $C_n = A_n B$ is a monotonically non-increasing sequence of events.

$$\therefore \quad P(\underset{n\to\infty}{Lt}\ C_n) = \underset{n\to\infty}{Lt}\ P(C_n). \tag{6.2.15}$$

Now $\underset{n\to\infty}{Lt}\ C_n = \prod_{n=1}^{\infty} C_n$. If the intersection is not empty, then there exists at least one point ω_1 (say) such that $\omega_1 \in C_n$ for all n, *i.e.*, $\omega_1 \in A_n B$ for all n.

So $\omega_1 \in A_n$ for all n.

$\therefore \quad a < X(\omega_1) \leq a + \frac{1}{n}$ for all n,

i.e., $a < c_1 \leq a + \frac{1}{n}$ for all n, where $X(\omega_1) = c_1$.

Now $c_1 - a > 0$. Then by Archimedean property, there exists a positive integer k such that

$$k(c_1 - a) > 1, \text{ i.e., } c_1 > a + \frac{1}{k}.$$

But $c_1 < a + \frac{1}{n}$ for every positive integer n. Hence we arrive at a contradiction. This proves that $\prod_{n=1}^{\infty} C_n = O$, the impossible event.

Then $P(O) = \underset{n\to\infty}{Lt}\ P(C_n)$ by (6.2.15)

$$= \underset{n\to\infty}{Lt}\ P(A_n B)$$

$$= \underset{n\to\infty}{Lt}\ P(a < X \leqslant a + \frac{1}{n},\ -\infty < Y \leqslant c)$$

$$= \underset{n\to\infty}{Lt} \left\{F(a+\frac{1}{n}, c) - F(a, c)\right\} \text{ by (6.2.6).}$$

So $$\underset{n\to\infty}{Lt} \left\{F(a+\frac{1}{n}, c) - F(a, c)\right\} = 0 \quad (6.2.16)$$

Now $F(x, y)$ being monotonically non-decreasing, in both the variables x and y, $\underset{x\to a+0}{Lt}\ F(x, y)$ exists finitely and

$$\underset{n\to\infty}{Lt}\ F\left(a + \frac{1}{n}, c\right) = \underset{x\to a+0}{Lt}\ F(x, c) = F(a+0, c).$$

Hence, by (6.2.16),

$$0 = F(a+0, c) - F(a, c),$$

i.e., $F(a+0, c) = F(a, c)$.

It can be similarly shown that $F(a, c+0) = F(a, c)$.

VI. Using (6.2.4) we can show that

$$F(b, d) + F(b-0, c) - F(b-0, d) - F(b, c)$$
$$= P(X = b, c < Y \leqslant d). \quad (6.2.17)$$

$$F(b, d) + F(a, d-0) - F(a, d) - F(b, d-0)$$
$$= P(a < X \leqslant b, Y = d). \quad (6.2.18)$$

$$F(b, d) + F(b-0, d-0) - F(b-0, d) - F(b, d-0)$$
$$= P(X = b, Y = d). \quad (6.2.19)$$

We state below (without proof) a set of conditions in which a given function of two variables will be a two-dimensional distribution function :

A real valued function $F(x, y)$ of two real variables x and y is a possible distribution function corresponding to a two dimensional random variable with respect to a suitable probability space if

(i) $F(x, y)$ is monotonically non-decreasing in both the variables and continuous to the right with respect to both the variables.

(ii) $F(-\infty, y) = F(x, -\infty) = 0$, $F(\infty, \infty) = 1$,

(iii) For every pair of points (x_1, y_1) and (x_2, y_2) with $x_1 < x_2$, $y_1 < y_2$,

$$F(x_2, y_2) + F(x_1, y_1) - F(x_2, y_1) - F(x_1, y_2) \geqslant 0.$$

It appears that a function of two variables requires an additional condition (*iii*) in order that it may be a possible distribution function. In fact properties (*i*) and (*ii*) are not sufficient conditions for a function of two variables to be a possible distribution function and this is evident from the following example :

Let $F(x, y)$ be a function of two variables defined by

$$F(x, y) = \begin{cases} 0, \text{ for } & x + y < 1 \\ 1, \text{ elsewhere.} \end{cases}$$

Here $F(x, y)$ satisfies properties (*i*) and (*ii*). But $F(x, y)$ is not a distribution function, since if it is the distribution function of a two-dimensional random variable (X, Y), then

$$P(\tfrac{1}{4} < X \leqslant 2, \tfrac{1}{4} < Y \leqslant 2)$$
$$= F(2, 2) + F(\tfrac{1}{4}, \tfrac{1}{4}) - F(2, \tfrac{1}{4}) - F(\tfrac{1}{4}, 2)$$
$$= 1 + 0 - 1 - 1$$
$$= -1 < 0$$

which is impossible and we see that $F(x, y)$ does not satisfy condition (*iii*).

6.3. Marginal Distributions.

Given the distribution of a two-dimensional random variable (X, Y), we now discuss how to find the individual distributions of

the random variables X and Y, called the *marginal distributions* of X and Y. Let $F(x, y)$ be the given distribution function of a two-dimensional random variable (X, Y). Then the marginal distributions of X and Y are respectively determined by the corresponding *marginal distribution functions* $F_X(x)$ and $F_Y(y)$ respectively, defined by

$$F_X(x) = P(-\infty < X \leq x),\ F_Y(y) = P(-\infty < Y \leq y).$$

Now from $F(x, y) = P(-\infty < X \leq x, -\infty < Y \leq y)$, (6.3.1)

proceeding to the limit $y \to \infty$ and keeping x fixed, we find that

$$F(x, \infty) = P(-\infty < X \leq x, -\infty < Y < \infty)$$
$$= P\{(-\infty < X \leq x) S\}, \text{ where } S \text{ is the certain event}$$
$$= P(-\infty < X \leq x) = F_X(x). \quad (6.3.2)$$

Similarly, by taking limit $x \to \infty$ in both sides of (6.3.1), keeping y fixed, it can be shown that

$$F(\infty, y) = P(-\infty < Y \leq y) = F_Y(y). \quad (6.3.3)$$

In (6.3.2) and (6.3.3), the distribution functions giving the marginal distributions of X and Y, *i.e.*, the individual distributions of X and Y, are obtained in terms of the joint distribution function of X and Y. Thus knowing the distribution function of the two-dimensional random variable (X, Y), we can find the marginal distributions of X and Y.

In general, if $F(x_1, x_2, \ldots, x_n)$ be the distribution function of an n-dimensional random variable $(X_1, X_2, \ldots, X_n)$, then the k-dimensional $(1 \leq k \leq n-1)$ marginal distribution function of the random variable $(X_{i_1}, X_{i_2}, \ldots, X_{i_k})$, where $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, is given by

$$\underset{\substack{x_i \to \infty \\ x_j \to \infty \\ \cdots\cdots \\ x_l \to \infty \\ (i, j, \ldots, l) \in \{1, 2, \ldots, n\} - \{i_1, i_2, \ldots, i_k\}}}{Lt} F(x_1, x_2, \ldots, x_n)$$
$$= F_{X_{i_1}, X_{i_2}, \ldots, X_{i_k}}(\infty, \ldots, \infty, x_{i_1}, \infty, \ldots, \infty, x_{i_2}, \infty, \ldots, x_{i_3}, \infty \ldots \infty, x_{i_k}, \infty, \ldots, \infty). \quad (6.3.4)$$

In particular, if $F(x, y, z)$ be the distribution function of the three dimensional random variable (X, Y, Z), then the marginal distribution functions of X, Y and Z are respectively

$$F_X(x) = F(x, \infty, \infty),\ F_Y(y) = F(\infty, y, \infty),\ F_Z(z) = F(\infty, \infty, z);$$

the marginal distribution function of the two dimensional random variable (X, Y) is $F_{X,Y}(x, y) = F(x, y, \infty)$, etc.

6.4. Independent Random Variables.

We have explained the notion of independence of two events in Chapter III and the independence of two random experiments in Chapter IV respectively. We now explain the concept of *independence of several random variables.*

Two random variables X, Y defined on the same event space S, are said to be independent if for any x, y the events $(-\infty < X \leq x)$ and $(-\infty < Y \leq y)$ are independent, *i.e.*,

$$P(-\infty < X \leq x, -\infty < Y \leq y) = P(-\infty < X \leq x)\, P(-\infty < Y \leq y)$$

or, $F(x, y) = F_X(x) F_Y(y)$ for all $x, y \in R$, (6.4.1)

where $F(x, y)$ is the joint distribution function of X and Y and $F_X(x)$, $F_Y(y)$ are the marginal distribution functions of X and Y respectively.

In general, if $F(x_1, x_2, \ldots, x_n)$ be the distribution function of the n-dimensional random variable $(X_1, X_2, \ldots, X_n)$, then the random variables $X_1, X_2, \ldots, X_n$ are said to be mutually independent or independent if

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad (6.4.2)$$

where $F_{X_i}(x_i)$ is the marginal distribution function of the random variable X_i $(i = 1, 2, \ldots, n)$.

Again the random variables $X_1, X_2, \ldots, X_n$ are said to be pairwise independent if

$$F_{X_i X_j}(x_i, x_j) = F_{X_i}(x_i)\, F_{X_j}(x_j) \quad (6.4.3)$$

for every pair of i and j, where $i \neq j$ and $i, j = 1, 2, \ldots, n$.

Finally, the random variables (X, Y) and Z are said to be independent if $F(x, y, z)=F_{X,Y}(x, y)\,F_Z(z)$, where $F(x, y, z)$ is the distribution function of the three dimensional random variable (X, Y, Z) and $F_{X,Y}(x, y)=F(x, y, \infty)$ and $F_Z(z)=F(\infty, \infty, z)$ are the marginal distribution functions of the random variables (X, Y) and Z respectively.

Theorem 6.4.1. *If X, Y, Z are mutually independent, then (a) X, Y, Z are pairwise independent and (b) (X, Y) and Z are independent.*

Proof: Let $F(x, y, z)$ be the distribution function of the three-dimensional random variable (X, Y, Z). Then if $F_{X,Y}(x, y)$, $F_X(x)$, $F_Y(y)$ and $F_Z(z)$ be the marginal distribution functions of (X, Y), X, Y and Z respectively,

$$F_{X,Y}(x, y)=F(x, y, \infty)=F_X(x)\,F_Y(y)\,F_Z(\infty)$$

since X, Y and Z are mutually independent

$$=F_X(x)F_Y(y), \quad (\because F_Z(\infty)=1) \tag{6.4.4}$$

which shows that X and Y are independent. Similarly Y and Z, Z and X are independent.

From (6.4.4), we get

$$F_Z(z)\,F_{X,Y}(x, y)=F_X(x)F_Y(y)F_Z(z)=F(x, y, z)$$

($\because$ X, Y and Z are mutually independent)

and hence (X, Y) and Z are independent.

In general, it can be easily seen that if $X_1, X_2, \ldots, X_n$ are mutually independent, then every subcollection $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ of $X_1, X_2, \ldots, X_n$ is also independent. We now prove few theorems on independence of two random variables.

Theorem 6.4.2. *A necessary and sufficient condition for two random variables X and Y to be independent is that, the joint distribution function $F(x, y)$ can be written as*

$$F(x, y)=\phi(x)\,\psi(y) \tag{6.4.5}$$

where $\phi(x)$ is a function of x only and $\psi(y)$ is a function of y only.

Proof: Let X and Y be independent.

Then $F(x, y)=F_X(x)\,F_Y(y)$, where $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions of X and Y respectively.

Taking $F_X(x)=\phi(x)$ and $F_Y(y)=\phi(y)$, we see that the condition is necessary.

Conversely, let $F(x, y)=\phi(x)\,\psi(y)$, where $\phi(x)$ is a function of x and $\psi(y)$ is a function of y only.

Then proceeding to the limit $x\to\infty$, we get

$$F(\infty, y)=\phi(\infty)\,\psi(y). \tag{6.4.6}$$

Also proceeding to the limit $y\to\infty$, we get

$$F(x, \infty)=\phi(x)\,\psi(\infty). \tag{6.4.7}$$

From (6.4.6) and (6.4.7), we get

$$F(x, \infty)\,F(\infty, y)=\phi(x)\,\psi(y)\,\phi(\infty)\,\psi(\infty).$$

Now from $F(\infty, \infty)=1$, we get $\phi(\infty)\,\psi(\infty)=1$.

$$\therefore \quad F(x, \infty)\,F(\infty, y)=\phi(x)\,\psi(y)$$

$$\text{or,} \quad F_X(x)\,F_Y(y)=F(x, y),$$

which proves that X and Y are independent.

Note. Thus if we want to show that two random variables X and Y are independent, it is sufficient to prove that the joint distribution function can be expressed as the product of two functions, one of which is a function of x only and the other a function of y only.

We now state, without proof, the corresponding theorem for n-dimensional joint distribution.

Theorem 6.4.3. *A necessary and sufficient condition for n random variables $X_1, X_2, \ldots, X_n$ to be mutually independent is that the distribution function $F(x_1, x_2, \ldots, x_n)$ of the n-dimensional random variable $(X_1, X_2, \ldots, X_n)$ can be written as*

$$F(x_1, x_2, \ldots, x_n)=\phi_1(x_1)\,\phi_2(x_2)\ldots\ldots\phi_n(x_n) \tag{6.4.8}$$

where $\phi_i(x_i)$ is a function of x_i only.

Theorem 6.4.4. *If X and Y are independent random variables, then*

$$P(a<X\leq b,\ c<Y\leq d)=P(a<X\leq b)\,P(c<Y\leq d) \tag{6.4.9}$$

Proof: Since X and Y are independent,

$$F(x, y) = F_X(x)\, F_Y(y) \text{ for all } x, y,$$

where $F(x, y)$ is the joint distribution function of X and Y, $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions of X and Y respectively.

Now $P(a < X \leqslant b, c < Y \leqslant d)$

$$= F(b, d) + F(a, c) - F(a, d) - F(b, c)$$

$$= F_X(b)\, F_Y(d) + F_X(a)\, F_Y(c) - F_X(a)\, F_Y(d) - F_X(b)\, F_Y(c),$$

since X and Y are independent

$$= \{F_Y(d) - F_Y(c)\}\{F_X(b) - F_X(a)\}$$

$$= P(c < Y \leqslant d)\, P(a < X \leqslant b).$$

The corresponding theorem for n-dimensional random variable $(X_1, X_2, \ldots, X_n)$ is stated below without proof.

Theorem 6.4.5. *If the random variables* $X_1, X_2, \ldots, X_n$ *are mutually independent, then*

$$P(a_1 < X_1 \leqslant b_1, a_2 < X_2 \leqslant b_2, \ldots, a_n < X_n \leqslant b_n)$$

$$= P(a_1 < X_1 \leqslant b_1)\, P(a_2 < X_2 \leqslant b_2) \ldots\ldots P(a_n < X_n \leqslant b_n). \quad (6.4.10)$$

Theorem 6.4.6. *If X and Y are independent random variables, then*

$$P(X = b, Y = d) = P(X = b)\, P(Y = d). \quad (6.4.11)$$

Proof: Since X and Y are independent random variables, we have by (6.4.9),

$$P(b - h < X \leqslant b, d - k < Y \leqslant d) = P(b - h < X \leqslant b)\, P(d - k < Y \leqslant d)$$

$$(h > 0, k > 0).$$

Now proceeding to the limit $h \to 0+$, $k \to 0+$, we get

$$P(X = b, Y = d) = P(X = b)\, P(Y = d).$$

The corresponding theorem for n-dimensional random variable $(X_1, X_2, \ldots, X_n)$ is stated below (without proof):

Theorem 6.4.7. *If the random variables* $X_1, X_2, \ldots, X_n$ *be mutually independent, then*

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

$$= P(X_1 = x_1)\, P(X_2 = x_2) \ldots P(X_n = x_n) \quad (6.4.12)$$

Probability Mass

Let $F(x, y)$ be the distribution function of a two-dimensional random variable (X, Y). Consider a mass distribution over the entire xy plane, the total mass being unity. Let the distribution of such a unit mass, which may vary from point to point, be such that the total mass distributed over the region $\{(x, y) : -\infty < x \leqslant a, -\infty < y \leqslant b\}$ is equal to $F(a, b)$. Then the probability of the event $(-\infty < X \leqslant a, -\infty < Y \leqslant b)$ can be interpreted as the mass distributed over the region $\{(x, y) : -\infty < x \leqslant a, -\infty < y \leqslant b\}$. Then such a mass distributed over the entire plane is called the probability mass of the two-dimensional random variable (X, Y).

Now $\{(-\infty < X \leqslant b, -\infty < Y \leqslant d) - (-\infty < X \leqslant b, -\infty < Y \leqslant c)\}$

$$-\{(-\infty < X \leqslant a, -\infty < Y \leqslant d) - (-\infty < X \leqslant a, -\infty < Y \leqslant c)\}$$

where $a < b$, $c < d$.

$$= (-\infty < X \leqslant b, c < Y \leqslant d) - (-\infty < X \leqslant a, c < Y \leqslant d)$$

$$= (a < X \leqslant b, c < Y \leqslant d).$$

Then by the aforesaid analogy of probability with the mass, we see that

probability mass in $\{(x, y) : a < x \leqslant b, c < y \leqslant d\}$

$=$ probability mass in $\{(x, y) : -\infty < x \leqslant b, -\infty < y \leqslant d\}$.

$-$ probability mass in $\{(x, y) : -\infty < x \leqslant b, -\infty < y \leqslant c\}$

$-$ probability mass in $\{(x, y) : -\infty < x \leqslant a, -\infty < y \leqslant d\}$

$+$ probability mass in $\{(x, y) : -\infty < x \leqslant a, -\infty < y \leqslant c\}$

$$= F(b, d) - F(b, c) - F(a, d) + F(a, c).$$

Hence we get

$$F(b, d) - F(b, c) - F(a, d) + F(a, c)$$

$=$ probability mass in $\{(x, y) : a < x \leqslant b, c < y \leqslant d\}$.

But left hand side is equal to $P(a < X \leqslant b, c < Y \leqslant d)$.

Thus the *probability mass* distributed in any finite rectangular region is equal to the probability that the random variable (X, Y) lies in that region.

6.5. Discrete Random Variable In Two Dimensions.

The two-dimensional random variable (X, Y) is said to be *discrete* if the spectrum of (X, Y) is at most countable. Here we assume that the spectrum of (X, Y) is a countable subset S_1 of R^2. Here for every pair $(x_i, y_j) \in S_1$ $(i, j=0, \pm 1, \pm 2, \ldots)$,

where $\cdots<x_{-2}<x_{-1}<x_0<x_1<x_2<\cdots$

and $\cdots<y_{-2}<y_{-1}<y_0<y_1<y_2<\cdots$,

we define the positive numbers f_{ij}, given by

$$f_{ij}=P(X=x_i, Y=y_j). \tag{6.5.1}$$

The necessary condition to be satisfied by f_{ij} is that

$$\sum_{(i,j)\in B}\sum f_{ij}=1, \tag{6.5.2}$$

where $B=\{(i, j) : (x_i, y_j) \in S_1\}$.

The distribution function of the two-dimensional discrete random variable (X, Y) is then given by

$$F(x, y)=\sum_{(\alpha,\beta)}\sum_{\in B'} f_{\alpha\beta} \text{ if } x_i \leqslant x < x_{i+1}, y_j \leqslant y < y_{j+1}, \tag{6.5.3}$$

where $B'=\{(\alpha, \beta) : x_\alpha \leqslant x_i, y_\beta \leqslant y_j\}$. Then $F(x,y)$ is a step function in two-dimensions with steps of heights f_{ij} (>0) at the points (x_i, y_j) $(i, j=0, \pm 1, \pm 2, \ldots)$.

Marginal Distribution.

Let (X, Y) be a two-dimensional discrete random variable which takes pairs of values (x_i, y_j) with probabilities

$$f_{ij}=P(X=x_i, Y=y_j), \ (i, j=0, \pm 1, \pm 2, \ldots).$$

Now the event $(X=x_i)$ can materialise when any one of the following mutually exclusive events happens :

$\ldots, (X=x_i, Y=y_{-2}), (X=x_i, Y=y_{-1}), (X=x_i, Y=y_0), \ldots$

Hence, $P(X=x_i)=\sum_{j=-\infty}^{\infty} P(X=x_i, Y=y_j)=\sum_{j=-\infty}^{\infty} f_{ij}=f_{i\cdot}$

$$\text{where } f_{i\cdot}=\sum_{j=-\infty}^{\infty} f_{ij}, \tag{6.5.4}$$

which is also denoted as f_{xi}.

Also the marginal distribution function $F_X(x)$ is given by

$$F_X(x)=F(x, \infty)=\sum_{x_i<x}\sum_{j=-\infty}^{\infty} f_{ij}=\sum_{x_i\leqslant x} f_{i\cdot} \tag{6.5.5}$$

This shows that $F_X(x)$ is a step function having steps of heights $f_{i\cdot}$ at x_i $(i=0, \pm 1, \pm 2, \ldots)$. It then follows that x_i $(i=0, \pm 1, \pm 2, \ldots)$ are the points of spectrum of X and $P(X=x_i)=f_{i\cdot}$. In other words, if we take the sum of all the masses on the line $x=x_i$ (for a fixed i), then for different values of i, we get the marginal distribution of X.

Similarly y_j's $(j=0, \pm 1, \pm 2, \ldots)$ denote the points of the spectrum of the single random variable Y and the probability mass $f_{\cdot j}$ at y_j is given by

$$f_{\cdot j}=P(Y=y_j)=\sum_{i=-\infty}^{\infty} f_{ij}, \tag{6.5.6}$$

which is also denoted by f_{yj}.

The marginal distribution function $F_Y(y)$ of the random variable Y is given by

$$F_Y(y)=F(\infty, y)=\sum_{i=-\infty}^{\infty}\sum_{y_j\leqslant y} f_{ij}=\sum_{y_j\leqslant y}\sum_{i=-\infty}^{\infty} f_{ij}$$

$$=\sum_{y_j\leqslant y} f_{\cdot j} \tag{6.5.7}$$

In other words, if we take the sum of all the masses on the line $y=y_j$ (for a fixed j), then for different values of j, we get the marginal distribution of Y.

It may be noted that

$$\sum_{i=-\infty}^{\infty} f_{i\cdot}=\sum_{i=-\infty}^{\infty}\left(\sum_{j=-\infty}^{\infty} f_{ij}\right)=1 \tag{6.5.8}$$

$$\text{and } \sum_{j=-\infty}^{\infty} f_{\cdot j}=\sum_{j=-\infty}^{\infty}\left(\sum_{i=-\infty}^{\infty} f_{ij}\right)=1. \tag{6.5.9}$$

Theorem 6.5.1. *A necessary and sufficient condition for two discrete random variables X and Y to be independent is that*

$$f_{ij}=f_{i\cdot}f_{\cdot j}(\text{or } f_{ij}=f_{xi}f_{yj}) \qquad (6.5.10)$$

where $f_{ij}=P(X=x_i,\ Y=y_j)$, (x_i, y_j) *being a spectrum point of the two-dimensional randon variable* (X, Y) *and* $f_{i\cdot}$ (*or* f_{xi}) *and* $f_{\cdot j}$ (*or* f_{yj}) *are the probability masses at* x_i *and* y_j *respectively of the corresponding marginal distributions of X and Y.*

Proof: Let X and Y be independent. Then by (6.4.11)

$$P(X=x_i,\ Y=y_j)=P(X=x_i)\ P(Y=y_j)$$

$$\therefore\ f_{ij}=f_{i\cdot}f_{\cdot j},$$

where $f_{i\cdot}=\sum_{j=-\infty}^{\infty} f_{ij}$ and $f_{\cdot j}=\sum_{i=-\infty}^{\infty} f_{ij}$, and hence the condition is necessary.

Conversely, let (6.5.10) hold. Then

$$F(x, y)=\sum_{(\alpha, \beta)\in B'}\sum f_{\alpha\beta}, \text{ if } x_i\leqslant x<x_{i+1},\ y_j\leqslant y<y_j+1,$$

where $B'=\{(\alpha, \beta) : x_\alpha\leqslant x_i,\ y_\beta\leqslant y_j\}$

$$=\sum_{(\alpha, \beta)\in B'}\sum f_{\alpha\cdot}f_{\cdot\beta}$$

$$=\Big(\sum_{x_\alpha\leqslant x_i} f_{\alpha\cdot}\Big)\Big(\sum_{y_\beta\leqslant y_j} f_{\cdot\beta}\Big)$$

$$=F_X(x)\ F_Y(y),$$

where F_X and F_Y are the marginal distribution functions of X and Y respectively. Since the above result holds for all i, j, we have $F(x, y)=F_X(x)\ F_Y(y)$ for all x, y. Hence X, Y are independent.

Finite Spectrum

We observe that the probability distribution of a two dimensional random variable (X, Y) is determined by the positive numbers $P(X=x_i, Y=y_j)=f_{ij}$, $(i, j=0, \pm 1, \pm 2, \ldots)$, where (x_i, y_j) is a point of the spectrum of (X, Y). If the spectrum be finite, then we can construct a table giving the values of f_{ij} for the corresponding pairs (x_i, y_j) and such a table is called **joint probability table.** So the joint distribution of X, Y having finite spectrum is determined by the joint probability table. We explain this by the following example.

The joint distribution of X, Y is given by the following table.

X \ Y	1	2	3	$P(X=x_i)$
0	·1	·3	·1	·5
2	·2	·1	·2	·5
$P(Y=y_j)$	·3	·4	·3	1

From the above table, we see that the marginal distribution of X is given by

$$x_i=i,\ i=0, 2,$$

with $f_{xi}=\sum_{j=1}^{3} f_{ij}$.

Then $f_{x0}=\cdot 5$, $f_{x2}=\cdot 5$.

Again the marginal distribution of Y is given by

$$y_j=j,\ j=1, 2, 3,$$

with $f_{yj}=f_{0j}+f_{2j}$.

Then we get $f_{y_1}=\cdot 3$, $f_{y_2}=\cdot 4$, $f_{y_3}=\cdot 3$.

Since $f_{x0}=\cdot 5$, $f_{y2}=\cdot 4$ and $f_{02}=\cdot 3$, so $f_{02}\neq f_{x0}f_{y2}$ and so the condition (6.5.10) is not fulfilled. Consequently the random variables X and Y are not independent.

6.6. Continuous Random Variable in two Dimensions.

Let X and Y be two random variables defined in the same event space S and $F(x, y)$ be the joint distribution function of X and Y.

The joint distribution of X and Y is said to be continuous, if $F(x, y)$ is continuous for all x and y and $\frac{\partial F}{\partial x}$, $\frac{\partial F}{\partial y}$, $\frac{\partial^2 F}{\partial x^2}$, $\frac{\partial^2 F}{\partial y^2}$, $\frac{\partial^2 F}{\partial y \partial x}$, $\frac{\partial^2 F}{\partial x \partial y}$ are all continuous in the whole xy-plane except that there may be a finite number of curves of discontinuities of these derivatives in any bounded region and for any $a, b, c, d \in R$, $\int_c^d \int_a^b \frac{\partial^2 F}{\partial x \partial y} dxdy$ is convergent, where $c<d$, $a<b$.

The *joint probability density function* $f(x, y)$ of the random variables X and Y is defined by

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y} \quad \cdots \quad (6.6.1)$$

provided the derivative $\frac{\partial^2 F}{\partial x \partial y}$ exists at the point (x, y).

$f(x, y)$ is also called the density function of the two-dimensional random variable (X, Y).

Theorem 6.6.1. $P(a<X\leqslant b, c<Y\leqslant d) = \int_c^d \int_a^b f(x,y)\, dxdy$, (6.6.2)

where $f(x, y)$ is the density function of the two-dimensional continuous random variable (X, Y).

Proof: We have, $\int_c^d \int_a^b f(x, y)\, dx\, dy = \int_c^d \int_a^b \frac{\partial^2 F}{\partial x \partial y} dx\, dy$

$$= \int_c^d \left\{ \int_a^b \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y}\right) dx \right\} dy$$

$$= \int_c^d \left\{ \frac{\partial F(b, y)}{\partial y} - \frac{\partial F(a, y)}{\partial y} \right\} dy$$

$$= F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

$$= P(a < X \leqslant b, c < Y \leqslant d).$$

Hence the theorem.

Theorem 6.6.2. $F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)\, dx\, dy$ (6.6.3)

where $F(x, y)$ and $f(x, y)$ are the distribution function and the density function of the two dimensional continuous random variable (X, Y) respectively.

Proof: Putting $b=x$, $d=y$ in (6.6.2), we get

$$P(a < X \leqslant x,\ c < Y \leqslant y) = \int_c^y \int_a^x f(x, y)\, dx dy$$

or, $F(x, y) - F(a, y) - F(x, c) + F(a, c) = \int_c^y \int_a^x f(x, y)\, dx\, dy.$

Now proceeding to the limit $a \to -\infty$, $c \to -\infty$, we get,

$$F(x, y) - \underset{a\to-\infty}{Lt} F(a, y) - \underset{c\to-\infty}{Lt} F(x, c) + \underset{c\to-\infty}{Lt}\ \underset{a\to-\infty}{Lt} F(a, c)$$

$$= \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)\, dx\, dy.$$

$$\therefore\ F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y)\, dx\, dy,$$

since all the limits vanish by (6.2.7).

Theorem 6.6.3. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$ (6.6.4)

where $f(x, y)$ is the density function of the two-dimensional continuous random variable (X, Y).

Proof: Proceeding to the limit $x \to \infty$, $y \to \infty$, we get from (6.6.3).

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

Theorem 6.6.4. *$f(x, y) \geqslant 0$ for all x, y, where $f(x, y)$ is the density function of the two-dimensional continuous random variable (X, Y).* (6.6.5)

Proof: Since $F(x, y)$ is monotonically increasing in both the variables x any y, $\frac{\partial^2 F}{\partial x \partial y} \geqslant 0$ *i.e.,* $f(x, y) \geqslant 0$ for all x, y.

Remark. A function $f(x, y)$ of two variables must satisfy the conditions (6.6.4) and (6.6.5) in order to be a possible joint density function of a two-dimensional continuous distribution.

Alternative definition of a two-dimensional continuous distribution.

The joint distribution of X and Y is said to be continuous if there exists a real valued non-negative function $f(x, y)$ such that $F(x, y)=\int_{-\infty}^{y}\left\{\int_{-\infty}^{x} f(u, v)\, du\right\} dv$ for all $x, y \in R$, where $F(x, y)$ is the distribution function of the two-dimensional random variable (x, y).

Here we find that at a point (x, y) of continuity of f, $\frac{\partial^2 F}{\partial x \partial y}=f(x, y)$. Further we observe that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy=1$.

This function f is called a joint probability density function of X and Y. We observe that the joint probability density function of X and Y defined in two ways determines the same distribution function $F(x, y)$.

Probability Differential (Two Dimensions) :

Let X and Y be two continuous random variables defined on the same event space S and let $F(x, y)$ be the distribution function of two-dimensional continuous random variable (X, Y). We assume that the second order partial derivatives of $F(x, y)$ exist.

$$\text{Now } \underset{\delta x \to 0}{Lt} \frac{1}{\delta x}\left[\underset{\delta y \to 0}{Lt}\left\{\frac{F(x+\delta x, y+\delta y)-F(x+\delta x, y)}{\delta y} - \frac{F(x, y+\delta y)-F(x, y)}{\delta y}\right\}\right]$$

$$= \underset{\delta x \to 0}{Lt} \frac{1}{\delta x}\left\{\frac{\partial F(x+\delta x, y)}{\partial y} - \frac{\partial F(x, y)}{\partial y}\right\}$$

$$= \underset{\delta x \to 0}{Lt} \frac{\frac{\partial F(x+\delta x, y)}{\partial y} - \frac{\partial F(x, y)}{\partial y}}{\delta x}$$

$$= \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Assuming that the above repeated limit is equal to the corresponding double limit, we get

$$\underset{\substack{\delta x \to 0 \\ \delta y \to 0}}{Lt} \frac{F(x+\delta x, y+\delta y)-F(x+\delta x, y)-F(x, y+\delta y)+F(x, y)}{\delta x\, \delta y}$$

$$= \frac{\partial^2 F(x, y)}{\partial x\, \partial y}$$

$$\text{or, } \underset{\substack{\delta x \to 0 \\ \delta y \to 0}}{Lt} \frac{P(x < X \leqslant x+\delta x, y < Y \leqslant y+\delta y)}{\delta x\, \delta y}=f(x, y),$$

where $f(x, y)$ is the density function of the joint distribution of X and Y.

Now the differentials of the independent variables being equal to their increments, we get

$$\underset{\substack{dx \to 0 \\ dy \to 0}}{Lt} \frac{P(x < X \leqslant x+dx, y < Y \leqslant y+dy)}{dx\, dy}=f(x, y).$$

We use the expression $P(x<X \leqslant x+dx, y<Y \leqslant y+dy)$ in this limiting sense and so without ambiguity $P(x < X \leqslant x+dx, y < Y \leqslant y+dy)$ is defined to be equal to $f(x, y)\, dx\, dy$. Then

$$P(x < X \leqslant x+dx, y < Y \leqslant y+dx)=f(x, y)\, dx\, dy. \qquad (6.6.6)$$

Probability Mass (Continuous Distribution).

In case of continuous two-dimensional random variable (X, Y), from the joint probability density function $f(x, y)$, we can determine the probability mass in any given region. If Q be a specified subset of the spectrum of the two-dimensional random variable (X, Y), then making an analogy with (6.6.2) we can write (formal proof is omitted),

$$P\{(X, Y) \in Q\} = \iint_Q f(x, y)\, dx\, dy. \qquad (6.6.7)$$

provided $\{(X, Y) \in Q\}=\{\omega : (X(\omega), Y(\omega)) \in Q, \omega \in S\}$ is a member of Δ, where (S, Δ, P) is the probability space.

Marginal Distribution for a Continuous Joint Distribution.

Let $F(x, y)$ be the distribution function of the two-dimensional continuous random variable (X, Y) and let $F_X(x)$ and $F_Y(y)$ be respectively the marginal distribution functions of X and Y. We have

$$F_X(x)=F(x, \infty)=\int_{-\infty}^{\infty}\int_{-\infty}^{x} f(x, y)\, dx\, dy$$

$$=\int_{-\infty}^{x}\left\{\int_{-\infty}^{\infty} f(x, y)\, dy\right\} dx$$

where $f(x, y)$ is the density function of the two-dimensional random variable (X, Y). Now $\int_{-\infty}^{\infty} f(x, y)\, dy$ is a function of x only and we denote this function by $f_X(x)$.

$$\therefore \quad F_X(x)=\int_{-\infty}^{x} f_X(x)\, dx, \tag{6.6.8}$$

which shows from the theory of one dimensional distribution that

$$f_X(x)=\int_{-\infty}^{\infty} f(x, y)\, dy \tag{6.6.9}$$

is the one-dimensional density function of the random variable X and it is called the *marginal density function* of X.

Similarly it can be shown that, if $f_Y(y)$ be the marginal density function of Y, then the marginal distribution function of Y is given by

$$F_Y(y)=\int_{-\infty}^{y} f_Y(y)\, dy \tag{6.6.10}$$

where $$f_Y(y)=\int_{-\infty}^{\infty} f(x, y)\, dx. \tag{6.6.11}$$

Thus $$F'_X(x)=f_X(x)=\int_{-\infty}^{\infty} f(x, y)\, dy, \tag{6.6.12}$$

and $$F'_Y(y)=f_Y(y)=\int_{-\infty}^{\infty} f(x, y)\, dx, \tag{6.6.13}$$

at points of continuity of $f_X(x)$ and $f_Y(y)$ respectively.

Theorem 6.6.5: *A necessary and sufficient condition for two continuous random variables X and Y to be independent is that the joint density function $f(x, y)$ of X and Y can be represented as*

$$f(x, y)=f_X(x)\, f_Y(y) \tag{6.6.14}$$

where $f_X(x)$ and $f_Y(y)$ are the marginal density functions of X and Y respectively.

Proof: Let X and Y be independent. Then

$$F(x, y)=F_X(x)\, F_Y(y)$$

where $F(x, y)$ is the distribution function of the two-dimensional random variable (X, Y) and $F_X(x)$ and $F_Y(y)$ are the marginal distribution functions of X and Y respectively.

$$\therefore \quad \frac{\partial F}{\partial x}=F'_X(x)\, F_Y(y)$$

and $$\frac{\partial^2 F}{\partial y\, \partial x}=F'_X(x)\, F'_Y(y).$$

But $\dfrac{\partial^2 F}{\partial y\, \partial x}=f(x, y)$ and $F'_X(x)=f_X(x)$, $F'_Y(y)=f_Y(y)$.

$\therefore\ f(x, y)=f_X(x)\, f_Y(y)$, so that the condition (6.6.14) is necessary.

Conversely, let $f(x, y)=f_X(x) f_Y(y)$

$$\therefore \quad \int_{-\infty}^{y}\int_{-\infty}^{x} f(x, y)\, dx\, dy=\left\{\int_{-\infty}^{x} f_X(x)\, dx\right\}\left\{\int_{-\infty}^{y} f_Y(y)\, dy\right\}$$

or, $F(x, y)=F_X(x)\, F_Y(y)$, by (6.6.3), (6.6.8) and (6.6.10).

Hence, the random variables X and Y are independent.

6·7. Two Important Continuous Two-Dimensional Distributions.

1. Uniform Distribution.

If the density function of a two-dimensional continuous random variable (X, Y) be constant C (say) in a fixed bounded region Q of the xy-plane and zero elsewhere, then the distribution is called a *uniform distribution*. Let the area of the fixed region Q be denoted by R. Then the joint probability density function of the uniform distribution is given by

$$f(x,y)=C, \text{ for } (x,y)\in Q$$
$$=0, \text{ elsewhere.}$$

Since $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1$, we get $C=\frac{1}{R}$.

Hence the uniform distribution is given by

$$f(x,y)=\frac{1}{R}, \text{ when } (x,y)\in Q$$
$$=0, \text{ elsewhere.} \tag{6.7.1}$$

If Q be a rectangular region $\{(x,y): a<x<b, c<y<d\}$, then the joint probability density function of the corresponding uniform distribution, called the *rectangular distribution*, is given by

$$f(x,y)=\frac{1}{(b-a)(d-c)}, \text{ for } a<x<b, c<y<d$$
$$=0, \text{ elsewhere.} \tag{6.7.2}$$

If Q' be a subregion of Q having area R', then by (6.6.7),

$$P\{(X,Y)\in Q'\}=\iint_{Q'} f(x,y)\,dx\,dy=\iint_{Q'}\frac{1}{R}\,dx\,dy=\frac{R'}{R}. \tag{6.7.3}$$

The marginal distribution of X, corresponding to the rectangular distribution (6.7.2) is given by

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy$$

$$\text{or, } f_X(x)=\int_c^d \frac{1}{(b-a)(d-c)}\,dy$$
$$=\frac{1}{b-a}, \quad a<x<b.$$

$$\text{i.e., } f_X(x)=\frac{1}{b-a}, \text{ when } a<x<b$$
$$=0, \text{ elsewhere.} \tag{6.7.4}$$

Hence X is uniformly distributed in the interval (a, b).

Similarly the marginal distribution of Y, corresponding to the rectangular distribution (6.7.2) is given by

$$f_Y(y)=\frac{1}{d-c}, \quad c<y<d$$
$$=0, \text{ elsewhere,} \tag{6.7.5}$$

so that Y is uniformly distributed in the interval (c, d).

Here we see that $f(x,y)=f_X(x)\,f_Y(y)$ for all x, y.

So X, Y are independent.

II. Bivariate Normal Distribution.

A two-dimensional continuous random variable (X, Y) is said to be normally distributed, if its joint density function $f(x,y)$ is given by

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\times e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_x)^2}{\sigma_x^2}+\frac{(y-m_y)^2}{\sigma_y^2}-2\rho\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y}\right\}},$$

$$(-\infty<x<\infty, -\infty<y<\infty) \tag{6.7.6}$$

where m_x, m_y, $\sigma_x\,(>0)$, $\sigma_y\,(>0)$ and $\rho\,(-1<\rho<1)$ are the five parameters of the distribution.

The marginal distribution of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_x)^2}{\sigma_x^2}+\frac{(y-m_y)^2}{\sigma_y^2}-2\rho\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y}\right\}} dy$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \underset{\substack{B_2\to\infty \\ B_1\to-\infty}}{Lt} \int_{B_1}^{B_2} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_x)^2}{\sigma_x^2}+\frac{(y-m_y)^2}{\sigma_y^2}-2\rho\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y}\right\}} dy$$

$$= \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}} \times \underset{\substack{B_2\to\infty \\ B_1\to-\infty}}{Lt} \int_{\frac{B_1-m_y}{\sigma_y}}^{\frac{B_2-m_y}{\sigma_y}} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_x)^2}{\sigma_x^2}+v^2-2\rho\frac{x-m_x}{\sigma_x}v\right\}} dy$$

where $v = \dfrac{y-m_y}{\sigma_y}$

$$= \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}} \underset{\substack{B_2\to\infty \\ B_1\to-\infty}}{Lt} \int_{\frac{B_1-m_y}{\sigma_y}}^{\frac{B_2-m_y}{\sigma_y}} e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)} dv$$

where $u = \dfrac{x-m_x}{\sigma_x}$

$$= \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}} \underset{\substack{B_2\to\infty \\ B_1\to-\infty}}{Lt} \int_{\frac{B_1-m_y}{\sigma_y}}^{\frac{B_2-m_y}{\sigma_y}} e^{-\frac{(v-\rho u)^2+(1-\rho^2)u^2}{2(1-\rho^2)}} dv$$

or, $$f_X(x) = \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}\,\sigma_x} \cdot \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}} dv$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{u^2}{2}} \cdot 1 = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}},$$

since for a normal (m, σ) distribution,

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} = 1,$$

so that taking $m = \rho u$ and $\sigma = \sqrt{1-\rho^2}$ (>0),

we get $$\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}} dv = 1.$$

$$\therefore\ f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}, \quad -\infty < x < \infty. \tag{6.7.7}$$

i.e., X is normal (m_x, σ_x).

Similarly, the marginal distribution of Y is also normal (m_y, σ_y) with corresponding density function

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}, \quad -\infty < y < \infty. \tag{6.7.8}$$

Note 1. It is important to note that if $\rho = 0$, then

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left\{\frac{(x-m_x)^2}{\sigma_x^2}+\frac{(y-m_y)^2}{\sigma_y^2}\right\}}$$

$$= \left\{\frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}\right\} \times \left\{\frac{1}{\sqrt{2\pi}\,\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}\right\}$$

$$= f_X(x) f_Y(y).$$

and hence by (6.6.14) the two random variables X and Y are independent in this case.

2. The marginal distributions of both X and Y are independent of the parameter ρ.

6·8. Conditional Distribution.

Let X and Y be two random variables defined on the same event space S. We explain the notion of conditional distribution separately when the joint distribution is discrete and the same is continuous.

(a) *Joint distribution of the random variables X and Y is discrete.*

Let the distribution of the two dimensional random variable (X, Y) be determined by

$$P(X=x_i, Y=y_j)=f_{ij},$$

for all (x_i, y_j) belonging to the spectrum of (X, Y).

We consider the probability of the event $(X=x_i)$ on the hypothesis that the event $(Y=y_j)$ has occurred. Then the non-negative real numbers $f_{i|j}$ defined by

$$f_{i|j}=P(X=x_i \mid Y=y_j), \tag{6.8.1}$$

where y_j is a fixed element of the spectrum of Y and x_i runs over the spectrum of X, are said to determine the *conditional distribution* of X, on the hypothesis $Y=y_j$.

Now $f_{i|j}=P(X=x_i \mid Y=y_j)=\dfrac{P(X=x_i, Y=y_j)}{P(Y=y_j)}=\dfrac{f_{ij}}{f_{yj}}.$

$$\therefore\ f_{i|j}=\frac{f_{ij}}{f_{yj}}, \tag{6.8.2}$$

for all possible values of i, determine the conditional distribution of X on the hypothesis $Y=y_j$, where $P(Y=y_j)=f_{yj}$

Similarly, the non-negative real number $f_{j|i}$ given by,

$$f_{j|i}=\frac{f_{ij}}{f_{xi}}, \tag{6.8.3}$$

where $P(X=x_i)=f_{xi}$, for different values of j, determine the conditional distribution of Y on the hypothesis $X=x_i$.

From (6.8.2) and (6.8.3) we see that

$$f_{ij}=f_{xi}\,f_{j|i}=f_{yj}\,f_{i|j}. \tag{6.8.4}$$

Now comparing (6.5.10) and (6.8.4) we observe that the condition of independence of the two random variables X and Y is also equivalent to

$$f_{i|j}=f_{xi} \quad \text{or} \quad f_{j|i}=f_{yj}. \tag{6.8.5}$$

(b) *Joint distribution of the random variables X and Y is continuous.*

Let the distribution of the two-dimensional continuous random variable (X, Y) be given by the density function $f(x, y)$.

In this case, $P(X \leqslant x \mid Y=y)$ for any fixed y and $P(Y \leqslant y \mid X=x)$ for any fixed x are not defined, since here $P(X=x)=P(Y=y)=0$, X and Y being both continuous. So to define the conditional distribution, we proceed as follows:

We assume that $f(x, y)$ and the marginal density function $f_Y(y)$ are both continuous functions in the xy-plane. Then for any fixed y and for any $\epsilon > 0$,

$$P(-\infty < X \leqslant x \mid y-\epsilon < Y \leqslant y+\epsilon)$$

$$=\frac{P(-\infty < X \leqslant x,\ y-\epsilon < Y \leqslant y+\epsilon)}{P(y-\epsilon < Y \leqslant y+\epsilon)}$$

$$=\frac{\displaystyle\int_{-\infty}^{x}\left\{\int_{y-\epsilon}^{y+\epsilon} f(x, t)\,dt\right\}dx}{\displaystyle\int_{y-\epsilon}^{y+\epsilon} f_Y(t)\,dt}$$

$$=\frac{\displaystyle\int_{-\infty}^{x} \epsilon f(x, y+\theta_1\epsilon)\,dx}{\epsilon f_Y(y+\theta_2\epsilon)},\quad 0<|\theta_1|,|\theta_2|<1$$

applying Mean Value Theorem of Integral Calculus.

Then proceeding to the limit $\epsilon \to 0+$, we get

$$\underset{\epsilon\to 0+}{Lt}\ P(-\infty < X \leqslant x \mid y-\epsilon < Y \leqslant y+\epsilon)$$

$$=\frac{\displaystyle\int_{-\infty}^{x} f(x, y)\,dx}{f_Y(y)},$$

$f(x, y)$ and $f_Y(y)$ being both continuous.

Thus, if we define the conditional distribution function $F_X(x \mid y)$ of the random variable X on the hypothesis $Y=y$, as

$$F_X(x \mid y) = \underset{\epsilon \to 0+}{Lt} P(-\infty < X \leqslant x \mid y-\epsilon < Y < y+\epsilon), \tag{6.8.6}$$

then $$F_X(x \mid y) = \frac{\int_{-\infty}^{x} f(x, y)\, dx}{f_Y(y)} = \int_{-\infty}^{x} f_X(x \mid y)\, dx, \tag{6.8.7}$$

where $$f_X(x \mid y) = \frac{f(x, y)}{f_Y(y)}. \tag{6.8.8}$$

Now we note that for all x and for any fixed y, $\frac{f(x, y)}{f_Y(y)} \geqslant 0$,

and $$\frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x, y)\, dx = \frac{f_Y(y)}{f_Y(y)} = 1.$$

So $f_X(x \mid y) = \frac{f(x, y)}{f_Y(y)}$, when y is fixed, is a possible density function.

So (6.8.7) and (6.8.8) indicate that $f_X(x \mid y)$ can be looked upon as the probability density function of a one dimensional distribution. This distribution is called the conditional distribution of X, given $Y=y$, and $f_X(x \mid y)$ is called the conditional density function of X on the hypothesis $Y=y$, $F_X(x \mid y)$ being called the conditional distribution function of X on the hypothesis $Y=y$. Now denoting

$\underset{\epsilon \to 0+}{Lt} P(-\infty < X \leqslant x \mid y-\epsilon < Y < y+\epsilon)$ as $P(-\infty < X \leqslant x \mid Y=y)$

we find that

$$P(-\infty < X \leqslant x \mid Y=y) = \int_{-\infty}^{x} f_X(x \mid y)\, dx. \tag{6.8.9}$$

Similarly, the conditional distribution function $F_Y(y \mid x)$ of the random variable Y, on the hypothesis $X=x$, for a fixed x, is defined by

$$F_Y(y \mid x) = \int_{-\infty}^{y} f_Y(y \mid x)\, dy \tag{6.8.10}$$

where $$f_Y(y \mid x) = \frac{f(x, y)}{f_X(x)}, \tag{6.8.11}$$

is the conditional density function of Y, on the hypothesis $X=x$, $f_X(x)$ being the marginal density function of X. From (6.8.8) and (6.8.11) we get

$$f_X(x \mid y) f_Y(y) = f_Y(y \mid x) f_X(x) = f(x, y). \tag{6.8.12}$$

Now comparing (6.6.14) and (6.8.12), we observe that the condition for independence of two continuous random variables X and Y reduces to

$$f_X(x \mid y) = f_X(x) \quad \text{or} \quad f_Y(y \mid x) = f_Y(y). \tag{6.8.13}$$

Let us now compute the conditional probability $P(a < X \leqslant b \mid Y=y)$ defined as $\underset{\delta \to 0+}{Lt} P(a < X \leqslant b \mid y-\delta < Y \leqslant y+\delta)$, $\delta > 0$.

Now $P(a < X \leqslant b \mid y-\delta < Y < y+\delta)$, $\delta > 0$

$$= \frac{P(a < X \leqslant b, y-\delta < Y < y+\delta)}{P(y-\delta < Y < y+\delta)}$$

$$= \frac{\int_a^b \left\{ \int_{y-\delta}^{y+\delta} f(x, t)\, dt \right\} dx}{\int_{y-\delta}^{y+\delta} f(t)\, dt}$$

$$= \frac{\int_a^b f(x, y+\theta_3\delta)\, dx}{f_Y(y+\theta_4\delta)}, \quad 0 < |\theta_3|, |\theta_4| < 1.$$

Hence, proceeding to the limit $\delta \to 0+$, we get

$$P(a < X \leqslant b \mid Y=y) = \frac{\int_a^b f(x, y)\, dx}{f_Y(y)} = \int_a^b f_X(x \mid y)\, dx. \tag{6.8.14}$$

Similarly, $$P(c < Y \leqslant d \mid X=x) = \int_c^d f_Y(y \mid x)\, dy. \tag{6.8.15}$$

6.9. Transformation of continuous random variables in two-dimensions.

Let X, Y be two continuous random variables.

Let $$u = \phi(x, y), \quad v = \psi(x, y), \tag{6.9.1}$$

be two continuous functions having continuous first order partial derivatives and let $\frac{\partial(u, v)}{\partial(x, y)}$ be either positive or negative throughout D. Then the inverse functions

$$x=h(u, v),\ y=k(u, v) \tag{6.9.2}$$

exist uniquely.

We consider the random variables U, V given by $U=\phi(X, Y)$, $V=\psi(X, Y)$, where $U=\phi(X, Y)$ is a mapping

$\phi(X, Y): S \to R$ defined by

$$[\phi(X, Y)]\,\omega=\phi[X(\omega), Y(\omega)],$$

S being the event space and $\psi(X, Y)$ is defined similarly.

Let $f_{X,Y}(x, y)$ and $f_{U,V}(u, v)$ be the density functions of (X, Y) and (U, V) respectively. Also let $f_{X,Y}(x, y) > 0$ in a domain $D \subseteq R^2$.

Since any elementary area $dx\,dy$ about the point (x, y) in the xy-plane is transformed [by the transformation (6.9.1)] uniquely into the elementary area $\left|\frac{\partial(x,y)}{\partial(u, v)}\right| du\,dv$ about the point (u, v) in the uv-plane, the event $(x < X \leqslant x+dx,\ y < Y \leqslant y+dy)$ implies and is implied by the event $(u < U \leqslant u+du,\ v < V \leqslant v+dv)$.

$$\therefore\quad P(x<X\leqslant x+dx,\ y<Y\leqslant y+dy)$$
$$=P(u<U\leqslant u+du,\ v<V\leqslant v+dv)$$

or, $f_{U,V}(u, v)\,du\,dv=f_{X,Y}(x, y)\,dx\,dy$

$$=f_{X,Y}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| du\,dv.$$

$$\therefore\quad f_{U,V}(u, v)=f_{X,Y}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| \tag{6.9.3}$$

gives the density function of the two dimensional random variable (U, V), where the expression on the right hand side is expressed as a function of u, v (using (6.9.2)).

Obtaining thus the density function of the random variable (U, V), we can proceed to find the marginal density function of the random variables U and V as follows :

$$f_U(u)=\int_{-\infty}^{\infty} f_{U,V}(u, v)\,dv=\int_{-\infty}^{\infty} f_{X,Y}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| dv \tag{6.9.4}$$

and
$$f_V(v)=\int_{-\infty}^{\infty} f_{U,V}(u, v)\,du=\int_{-\infty}^{\infty} f_{X,Y}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| du. \tag{6.9.5}$$

Distribution of the sum of two continuous random variables. Let $f_{X,Y}(x, y)$ be the density function of the continuous random variable (X, Y).

Let $U=X+Y,\ V=X$.

In terms of real variables $u=x+y,\ v=x$.

Then $\frac{\partial(u, v)}{\partial(x, y)}=\begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix}=-1<0$ for all x, y.

If $f_U(u)$ be the marginal density function of U, then by (6.9.4).

$$f_U(u)=\int_{-\infty}^{\infty} f_{X,Y}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| dv$$
$$=\int_{-\infty}^{\infty} f_{X,Y}(v, u-v)\,dv. \tag{6.9.6}$$

Note. If X and Y are independent, then $f_{X,Y}(x, y)=f_X(x)f_Y(y)$ and so by (6.9.6) we get $f_U(u)=\int_{-\infty}^{\infty} f_X(v)f_Y(u-v)\,dv.$ (6.9.7)

Theorem 6.9.1. *Let $u=g_1(x)$, $v=g_2(y)$ be two continuously differentiable and strictly monotonic functions of x and y respectively. If the random variables X and Y are independent, then the random variables U and V, defined by $U=g_1(X)$ and $V=g_2(Y)$, are also independent.*

Proof : Since $g_1(x)$ and $g_2(y)$ are continuously differentiable and strictly monotonic functions, the distribution of the random variables U and V are determined respectively by the density function $f_U(u)$ and $f_V(v)$, where

$$f_U(u)=f_X(x)\left|\frac{dx}{du}\right| \text{ and } f_V(v)=f_Y(y)\left|\frac{dy}{dv}\right|.$$

$f_X(x)$ and $f_Y(y)$ being the density functions of X and Y respectively.

Now $\frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix}\frac{du}{dx} & 0\\ 0 & \frac{dv}{dy}\end{vmatrix}=\frac{du}{dx}\,\frac{dv}{dy}=g_1'(x)\,g_2'(y)>\text{ or }<0$

for all x, y, since the functions $g_1(x)$ and $g_2(y)$ are strictly monotonic. Now, if $f_{U,V}(u,v)$ and $f_{X,Y}(x,y)$ be the density functions of the two dimensional random variables (X, Y) and (U, V), then by (6.9.3),

$$f_{U,V}(u,v)=f_{X,Y}(x,y)\left|\frac{dx}{du}\right|\left|\frac{dy}{dv}\right|$$

$$=\left\{f_X(x)\left|\frac{dx}{du}\right|\right\}\left\{f_Y(y)\left|\frac{dy}{dv}\right|\right\},$$

since the random variables X and Y are independent.

$$=f_U(u)\,f_V(v).$$

This proves that U and V are independent.

Transformation of continuous random variables in n-dimensions.

We state below (without proof) the theorem which gives the rule for finding the probability density function of an n-dimensional random variable $(U_1, U_2, \ldots, U_n)$ obtained by a transformation of the type mentioned below from an n-dimensional random variable $(X_1, X_2, \ldots, X_n)$.

Theorem 6.9.2. *Let $(X_1, X_2, \ldots, X_n)$ be an n dimensional random variable of the continuous type with probability density function*

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n).$$

Let (*i*)

$$u_1=f_1(x_1, x_2, \ldots, x_n)$$
$$u_2=f_2(x_1, x_2, \ldots, x_n)$$
$$\ldots \quad \ldots \quad \ldots$$
$$u_n=f_n(x_1, x_2, \ldots, x_n)$$

give a bijective mapping of $R^n \to R^n$ i.e., there exists the inverse transformation

$$x_1=\phi_1(u_1, u_2, \ldots, u_n)$$
$$x_2=\phi_2(u_1, u_2, \ldots, u_n)$$
$$\ldots \quad \ldots \quad \ldots$$
$$x_n=\phi_n(u_1, u_2, \ldots, u_n)$$

defined over the corresponding domain,

(*ii*) *the given tranformation and its inverse are continuous,*

(*iii*) *the partial derivatives $\frac{\partial x_i}{\partial u_j}$ $(1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant n)$ are continuous,*

(*iv*) *the Jacobian* $J=\begin{vmatrix}\frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \cdots & \frac{\partial x_1}{\partial u_n}\\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} & \cdots & \frac{\partial x_2}{\partial u_n}\\ \cdots & \cdots & & \cdots\\ \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \cdots & \frac{\partial x_n}{\partial u_n}\end{vmatrix}\neq 0$

for all $(u_1, u_2, \ldots, u_n)$ in the corresponding domain.

Then the n-dimensional random varvable $(U_1, U_2, \ldots, U_n)$ defined by $U_i=f_i(X_1, X_2, \ldots, X_n)$, $i=1, 2, \ldots, n$, is continuous with probability density function $f_{U_1, U_2, \ldots, U_n}(u_1, u_2, \ldots, u_n)$ given by

$$f_{U_1, U_2, \ldots, U_n}(u_1, u_2, \ldots, u_n)$$
$$=f_{X_1, X_2, \ldots, X_n}(\phi_1(u_1, u_2, \ldots, u_n), \phi_2(u_1, u_2, \ldots, u_n), \ldots,$$
$$\phi_n(u_1, u_2, \ldots, u_n))\times|J|.$$

We now state the theorem (without proof) in n dimensions corresponding to the theorem 6.9.1.

Theorem 6.9.3. *If $X_1, X_2, \ldots, X_n$ are mutually independent continuous random variables and if*

$$g_1(x_1, x_2, \ldots, x_{k_1}),\ g_2(x_{k_1+1}, x_{k_1+2}, \ldots, x_{k_2}), \ldots,$$

$g_{m+1}(x_{k_m+1}, x_{k_m+2}, \ldots, x_n)$ be continuous functions of their arguments and if the corresponding inverse functions exist, then the random variables

$$g_1(X_1, X_2, \ldots, X_{k_1}),\ g_2(X_{k_1+1}, X_{k_1+2}, \ldots, X_{k_2}), \ldots,$$

$g_{m+1}(X_{k_m+1}, X_{k_m+2}, \ldots, X_n)$ are mutually independent.

6.10. Illustrative Examples.

Ex. 1. *A fair coin is tossed three times. Let X denote the number of heads in three tossings and Y denote the absolute difference between the number of heads and number of tails. Find the*

joint p.m.f. of (X, Y) and the marginal p.m.f. of X and Y. Are then two randome variables X and Y independent? Find the conditional p.m.f. of X, given $Y=1$.

The event space of the given random experiment consists of the following 8 out comes :

$(H, T, T), (H, T, H), (H, H, T), (H, H, H), (T, H, H), (T, T, T), (T, H, T), (T, T, H)$.

Hence, the spectrum of X and Y are given by

$$x_i = i \ (i=0, 1, 2, 3)$$
$$y_j = j \ (j=1, 3),$$

and the spectrum of the two dimensional random variable (X, Y) is

$$(x_i, y_j) = (i, j)$$
$$(i=0, 1, 2, 3 ; j=1, 3).$$

The p.m.f. of (X, Y) is given by the following table :

$X \backslash Y$	1	3	$P(X=x_i)$
0	0	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{3}{8}$	0	$\frac{3}{8}$
2	$\frac{3}{8}$	0	$\frac{3}{8}$
3	0	$\frac{1}{8}$	$\frac{1}{8}$
$P(Y=y_j)$	$\frac{3}{4}$	$\frac{1}{4}$	1

The above table is obtained from consideration of the number of district out comes favourable to the event $(X=i, Y=j)$, for example, the event $(X=2, Y=1)$ contains 3 distirict outcomes $(H, T, H), (H, H, T), (T, H, H)$ and assuming that all the simple events are equally likely, we find that $P(X=2, Y=1)=\frac{3}{8}$. The row totals shown in column gives the marginal p.m.f. of X and the column totals shown in the rows represents the marginal p.m.f. of Y.

Thus the distribution of the two-dimensional random variable (X, Y) can be expressed as follows :

The distribution of (X, Y) is given by probability masses $p_{ij}=P(X=i, Y=j)$,

where $p_{11}=p_{21}=\frac{3}{8}$, $p_{03}=p_{33}=\frac{1}{8}$, $p_{01}=p_{31}=p_{13}=p_{23}=0$.

The marginal distribution of X is as follows :

The spectrum of X is $x_i=i \ (i=0, 1, 2, 3)$ with probability masses $p_{x_i}=P(X=x_i)$, where $p_{x_0}=\frac{1}{8}$, $p_{x_1}=p_{x_2}=\frac{3}{8}$, $p_{x_3}=\frac{1}{8}$.

The marginal distribution of Y is given by

$y_j=j \ (j=1, 3)$ with p.m.f. $p_{y_j}=P(Y=y_j)$, where $p_{y_1}=\frac{3}{4}$, $p_{y_3}=\frac{1}{4}$.

Now $p_{13}=0$ and $p_{x_1}\cdot p_{y_3}=\frac{3}{32}$, so that $p_{13}\neq p_{x_1}\cdot p_{y_3}$ and hence X and Y are not independent.

Again the conditional p.m.f. $p_{i|1}$ of X on the hypothesis $Y=1$ is given by

$$p_{i|1}=\frac{P(X=i, Y=1)}{P(Y=1)}=\frac{p_{i1}}{\frac{3}{4}}.$$

$$\therefore \ p_{i|1}=\begin{cases} 0 & \text{for } i=0, 3 \\ \frac{1}{2} & \text{for } i=1, 2. \end{cases}$$

Ex. 2. *Consider the random experiment of throwing a pair of dice. Let X denote the number of sixes and Y denote the number of fives that turn up. Find the joint p. m. f. of the two-dimensional random variable (X, Y) and the marginal p. m. f. of X and Y.*

Find the probability

$$P(X+Y \geqslant 2).$$

Here the spectra of X and Y are given by

$$x_i=i \ (i=0, 1, 2) \text{ and } y_j=j \ (j=0, 1, 2)$$

respectively and the spectrum of the two-dimensional random variable (X, Y) is given by

$$(x_i, y_j)=(i, j), \ (i=0, 1, 2 ; j=0, 1, 2) \text{ with } P(X=i, Y=j)>0.$$

The p.m.f. of (X, Y) is given by the following table ;

$X \backslash Y$	0	1	2	$P(X=x_i)$
0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\frac{25}{36}$
1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\frac{10}{36}$
2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
$P(Y=y_j)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	1

The above table is constructed from consideration of the number of distinct outcomes favourable to the events of the type $(X=i, Y=j)$, for example, the event $(X=0, Y=1)$ contains 8 distinct outcomes (1, 5), (2, 5), (3, 5), (4, 5), (5, 4), (5, 3), (5, 2), (5, 1) and assuming that all the simple events are equally likely, we find that $P(X=0, Y=1)=\frac{8}{36}$. Thus the distribution of the two-dimensional random variable (X, Y) can be expressed as follows :

The spectrum of (X, Y) is given by

$(x_i, y_j)=(i, j)$, $(i=0, 1, 2, j=0, 1, 2)$ excluding the points (2, 1), (2, 2), (1, 2) with corresponding p.m.f. $p_{ij}=P(X=i, Y=j)$, where $p_{00}=\frac{16}{36}$, $p_{01}=p_{10}=\frac{8}{36}$, $p_{02}=p_{20}=\frac{1}{36}$, $p_{11}=\frac{2}{36}$.

From the row sum of the above table, the marginal distribution of X is as follows :

The spectrum of X is given by $x_i=i$ $(i=0, 1, 2)$ and p. m. f. is given by

$$p_{x_i}=P(X=i),\ (i=0, 1, 2), \text{ where } p_{x0}=\tfrac{25}{36},\ p_{x1}=\tfrac{10}{36},\ p_{x2}=\tfrac{1}{36}.$$

Similarly, the spectrum of Y is given by $y_j=j$ $(j=0, 1, 2)$ and p.m.f. is given by

$$p_{yj}=P(Y=j),\ (j=0, 1, 2), \text{ where } p_{y0}=\tfrac{25}{36},\ p_{y1}=\tfrac{10}{36},\ p_{y2}=\tfrac{1}{36}.$$

Finally, $P(X+Y \geqslant 2)=p_{11}+p_{02}+p_{20}=\frac{2}{36}+\frac{1}{36}+\frac{1}{36}=\frac{1}{9}$.

Ex. 3. *Two balls are drawn with replacement/without replacement from an urn containing 2 white, 2 black and 4 red balls. Let X_k $(k=1, 2)$ be the random variable taking values 1 or 0, according as the ball drawn on the kth draw is white or non-white. Find the joint p. m. f. of (X_1, X_2). Deduce the p. m. f. for the marginal distributions of X_1 and X_2. Find the conditional distribution of X_1 on the hypothesis $X_2=1$.*

Case I. With replacement.

The spectrum of the two-dimensional random variable (X_1, X_2) is

$$(x_i, y_j)=(i, j)\ (i=1, 0\ ;\ j=1, 0).$$

If $p_{ij}=P(X_1=i, X_2=j)$, then

$p_{00}=P(X_1=0, X_2=0)$

= probability of the event that both the balls drawn (with replacement) are non-white.

$=\frac{6}{8}\times\frac{6}{8}=\frac{9}{16}$.

Similarly, $p_{10}=P(X_1=1, X_2=0)=\frac{2}{8}\times\frac{6}{8}=\frac{3}{16}$,

$p_{01}=P(X_1=0, X_2=1)=\frac{6}{8}\times\frac{2}{8}=\frac{3}{16}$,

$p_{11}=P(X_1=1, X_2=1)=\frac{2}{8}\times\frac{2}{8}=\frac{1}{16}$.

The p.m.f. for the two-dimensional random variable (X_1, X_2) along with the marginal p.m.f. for X_1 and X_2 are shown by the following table :

$X_2 \backslash X_1$	0	1	$P(X_2=j)$
0	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{12}{16}$
1	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
$P(X_1=i)$	$\frac{12}{16}$	$\frac{4}{16}$	1

Hence the marginal distribution of X_1 is given by

$$X_1=i\ (i=0, 1)$$

with p.m.f. given by $p_{x_1 i}=P(X_1=i)$, where

$$p_{x_1 0}=\tfrac{3}{4},\ p_{x_1 1}=\tfrac{1}{4}.$$

and that of X_2 is

$$X_2 = j, \ (j = 0, 1)$$

with probability masses $p_{x,j} = P(X_2 = j)$, where

$$p_{x,0} = \tfrac{3}{4}, \ p_{x,1} = \tfrac{1}{4}.$$

Case II. Without replacement.

In this case, $p_{00} = P(X_1 = 0, \ X_2 = 0) = \frac{6}{8} \times \frac{5}{7} = \frac{15}{28}$,

$p_{01} = P(X_1 = 0, \ X_2 = 1) = \frac{6}{8} \times \frac{2}{7} = \frac{6}{28}$,

$p_{10} = P(X_1 = 1, \ X_2 = 0) = \frac{2}{8} \times \frac{6}{7} = \frac{6}{28}$,

$p_{11} = P(X_1 = 1, \ X_2 = 1) = \frac{2}{8} \times \frac{1}{7} = \frac{1}{28}$.

Hence the following table gives the p.m.f. for the two-dimensional random variable (X_1, X_2) along with the marginal distributions of X_1 and X_2 :

$X_2 \backslash X_1$	0	1	$P(X_2 = j)$
0	$\frac{15}{28}$	$\frac{6}{28}$	$\frac{21}{28}$
1	$\frac{6}{28}$	$\frac{1}{28}$	$\frac{7}{28}$
$P(X_1 = i)$	$\frac{21}{28}$	$\frac{7}{28}$	1

In case I, the conditional distribution of X_1 on the hypothesis $X_2 = 1$ is given by

$$p_{i|1} = \frac{p_{i1}}{p_{x,1}} = 4p_{i1},$$

$$\text{i.e., } p_{0|1} = 4p_{01} = \frac{3}{4}, \ p_{1|1} = 4p_{11} = \frac{1}{4}.$$

In Case II, the conditional distribution of X_1 on the hypothesis $X_2 = 1$ is given by

$$p_{i|1} = \frac{p_{i1}}{p_{x,1}} = \frac{28}{7} p_{i1}, \text{ so that}$$

$$p_{0|1} = \frac{28}{7} p_{01} = \frac{6}{7}, \ p_{1|1} = \frac{28}{7} p_{11} = \frac{1}{7}.$$

Ex. 4. Let $p = \frac{1}{2}$ be the probability of a female birth in a family of three children. Let X be a random variable denoting the number of female child in the first two births and Y be the number of female child in the last two births. Find the joint distribution of the two-dimensional random variable (X, Y), the marginal distributions of X and Y, $P(Y = 2 \mid X = 1)$.

The spectrum of the two-demensional random variable (X, Y) is $(x_i, y_j) = (i, j)$, $(i = 0, 1, 2 \,;\ j = 0, 1, 2)$ excluding $(0, 2)$ and $(2, 0)$.

The probability masses at different spectrum points along with marginal probability masses of X and Y are given by the following table :

$Y \backslash X$	0	1	2	$P(Y=j)$
0	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
$P(X=i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

The above table is obtained as follows :

The problem can be considered as Bernoulli trials with $n = 3$ and with probability of 'success' $p = \frac{1}{2}$, where 'success' represents a female birth. The corresponding event space contains 8 outcomes

(M, M, M), (M, F, M), (M, M, F), (M, F, F), (F, M, F), (F, F, M), (F, F, F), (F, M, M), where F and M represent 'a female birth' and 'a male birth' respectively.

Then the event $(X = 1, \ Y = 1) = \{(M, F, M), \ (F, M, F)\}$, and so $P(X = 1, Y = 1) = \frac{2}{8} = \frac{1}{4}$;

the event $(X = 1, \ Y = 0) = \{(F, M, M)\}$, and $P(X = 1, \ Y = 0) = \frac{1}{8}$; $(X = 0, Y = 2)$ is an impossible event and so on.

The conditional distribution of Y on the hypothesis $X=1$ is given by p. m. f. $p_{j|1}=\dfrac{p_{1j}}{p_{x1}}=2p_{1j}$.

$$\therefore\quad P(Y=2 \mid X=1)=2p_{12}=\tfrac{1}{2}.$$

Ex. 5. *A two-dimensional random variable (X, Y) has the spectrum $(x_i, y_j)=(i, j)$; $(i=0, 1, 2, 3\ ;\ j=1, 2, 3, 4)$ and the joint probabilities p_{ij} are given by*

$$p_{ij}=P(X=i, Y=j)=c(3i+4j), \text{ where } c \text{ is a constant.}$$

Find (a) *the value of c,*

(b) *the marginal distributions of X and Y,*

(c) $P(X>2, Y<3)$,

(d) $P(Y=2 \mid X=3)$.

Also examine whether X and Y are independent.

(a) From the necessary condition $\Sigma p_{ij}=1$, we get

$$c\sum_{i=0}^{3}\sum_{j=1}^{4}(3i+4j)=1$$

$$\text{or,}\quad c\sum_{i=0}^{3}(3i+4+3i+8+3i+12+3i+16)=1$$

$$\text{or,}\quad c\sum_{i=0}^{3}(12i+40)=1$$

$$\text{or,}\quad c(40+12+40+24+40+36+40)=1.$$

$$\therefore\quad c=\frac{1}{232}.$$

(b) The marginal distribution of X has the spectrum given by

$$x_i=i,\ (i=0, 1, 2, 3)$$

and p. m. f. is given by $p_{xi}=P(X=x_i)$, where

$$p_{xi}=\sum_{j=1}^{4}p_{ij}=\frac{1}{232}\sum_{j=1}^{4}(3i+4j)$$

$$=\frac{1}{232}(12i+40)=\frac{1}{58}(3i+10),\ i=0, 1, 2, 3.$$

The marginal distribution of Y has the spectrum given by

$$y_j=j \qquad (j=1, 2, 3, 4)$$

with probability masses $p_{yj}=P(Y=y_j)$, where

$$p_{yj}=\frac{1}{232}\sum_{i=0}^{3}(3i+4j)=\frac{1}{232}(18+16j)$$

$$=\frac{1}{116}(9+8j),\ j=1, 2, 3, 4.$$

(c)
$$P(X>2, Y<3)=\sum_{i\geqslant 2}\sum_{j\leqslant 3}p_{ij}$$

$$=\frac{1}{232}\sum_{i\geqslant 2}\sum_{j=1}^{3}(3i+4j)$$

$$=\frac{1}{232}\sum_{i\geqslant 2}(3i+4+3i+8+3i+12)$$

$$=\frac{1}{232}\sum_{i=2}^{3}(9i+24)$$

$$=\frac{1}{232}(18+24+27+24)=\frac{93}{232}.$$

(d)
$$P(Y=2 \mid X=3)=\frac{P(X=3, Y=2)}{P(X=3)}$$

$$=\frac{p_{32}}{p_{x3}}=\frac{\frac{1}{232}(9+8)}{\frac{1}{58}(9+10)}=\frac{17}{76}.$$

We have $p_{12}=\dfrac{1}{232}(3+8)=\dfrac{11}{232}$, $p_{x1}\,p_{y2}=\dfrac{13}{58}\times\dfrac{25}{116}$,

so that $p_{12}\neq p_{x1}\,p_{y2}$. Hence X and Y are not independent.

Ex. 6. *The joint distribution of (X, Y) is defined by*

$$P(X=0, Y=0)=P(X=0, Y=1)=P(X=1, Y=1)=\tfrac{1}{3}.$$

(i) *Find the distribution function of (X, Y) and the marginal distribution functions of X and Y.*

(ii) *Examine whether the point $P_1(-\frac{1}{2}, 0)$, $P_2(0, 1)$ are points of continuity of the joint distribution function $F(x, y)$.*

The joint distribution of (X, Y) is as follows :

The spectrum points are

$(x_i, y_j)=(i, j)$; $i=0, 1$ and $j=0, 1$ excluding $(1, 0)$

with $p. m. f.$ $p_{ij}=P(X=i, Y=j)$, where

$p_{00}=\frac{1}{3}, p_{01}=\frac{1}{3}, p_{11}=\frac{1}{3}$.

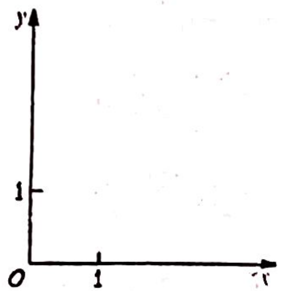


Fig. 6.10.1

The event space is shown in Fig. 6.10.1.

(i) The distribution function $F(x, y)$ is given by

$$F(x, y)=\sum_{i<0}\sum_{j<0} p_{ij}=0 \text{ for } x<0 \text{ or } y<0,$$

$$F(x, y)=\sum_{i\geq 0}\sum_{0\leq j<1} p_{ij}=p_{00}=\tfrac{1}{3}, \text{ for } 0\leq x<\infty,\ 0\leq y<1,$$

$$F(x, y)=\sum_{0\leq i<1}\sum_{1\leq j<\infty} p_{ij}=p_{00}+p_{01}=\tfrac{2}{3}, \text{ for } 0\leq x<1,\ 1\leq y<\infty$$

$$F(x, y)=\sum_{i\geq 1}\sum_{j\geq 1} p_{ij}=p_{00}+p_{01}+p_{11}=1, \text{ for } x\geq 1, y\geq 1.$$

$$\text{Thus, } F(x, y)=\begin{cases} 0, & \text{for } x<0 \text{ or } y<0 \\ \tfrac{1}{3}, & \text{for } 0\leq x<\infty,\ 0\leq y<1 \\ \tfrac{2}{3}, & \text{for } 0\leq x<1,\ 1\leq y<\infty \\ 1, & \text{for } x\geq 1, y\geq 1. \end{cases}$$

The $p. m. f.$ of the marginal distribution of X is

$p_{x0}=\frac{2}{3}, p_{x1}=\frac{1}{3}$,

the spectrum of the distribution being given by $x_i=i$, $i=0, 1$.

The following table gives the joint distribution of X and Y.

Y \ X	0	1	$P(Y=j)$
0	$\frac{1}{3}$	0	$\frac{1}{3}$
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
$P(X=i)$	$\frac{2}{3}$	$\frac{1}{3}$	1

The $p. m. f.$ of the marginal distribution of Y is

$p_{y0}=\frac{1}{3}, p_{y1}=\frac{2}{3}$, the spectrum of the distribution being given by

$y_j=j$, $j=0, 1$.

The distribution function $F_X(x)$ is given by $F_X(x)=\begin{cases} 0, & \text{for } x<0 \\ \frac{2}{3}, & \text{for } 0\leq x<1 \\ 1, & \text{for } x\geq 1. \end{cases}$

The distribution function $F_Y(y)$ is given by $F_Y(y)=\begin{cases} 0, & \text{for } y<0 \\ \frac{1}{3}, & \text{for } 0\leq y< \\ 1, & \text{for } y\geq 1. \end{cases}$

(ii) Now, $\underset{\substack{x\to -\frac{1}{2} \\ y\to 0}}{Lt} F(x, y)=0$, since $F(x, y=0$ if $x<0$ or $y<0$.

Again, $F(-\frac{1}{2}, 0)=0$. Hence $F(x, y)$ is continuous at $(-\frac{1}{2}, 0)$. $(0, 1)$ is a point of discontinuity of $F(x, y)$ because for $x<0$, $F(x, 1)=0$, while $F(0, 1)=\frac{2}{3}$.

Ex 7. *Raindrops fall at random on a square R with vertices $(1, 0)$, $(0, 1)$, $(-1, 0)$, $(0, -1)$. An outcome is the point (x, y) in R struck by a particular raindrop. Let $X(x, y)=x$, $Y(x, y)=y$ and assume (X, Y) has uniform distribution over R. Determine the joint and marginal distributions of X and Y. Are the random variables X and Y independent ?*

[C. H. (Math.) '81]

Here the two-dimensional random variable (X, Y) is uniform over the region R, a square region of area 2 square units, having vertices $(0, 1)$, $(-1, 0)$, $(0, -1)$ and $(1, 0)$ (Fig. 6.10.2). So the

probability density function of the joint distribution of X and Y is given by

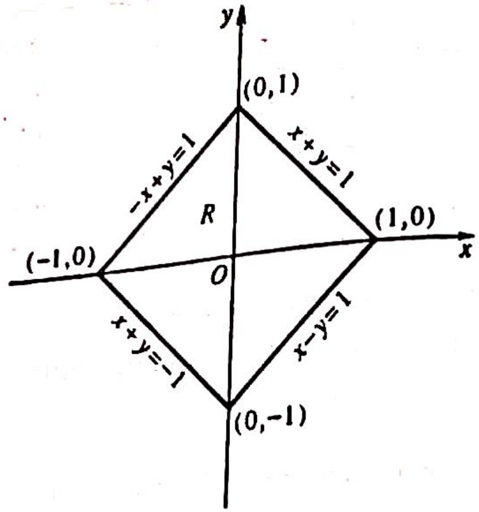


Fig. 6.10.2

$$f(x, y)=\begin{cases} \frac{1}{2}, & \text{if } (x, y) \in R \\ 0, & \text{elsewhere.} \end{cases}$$

Marginal density function $f_X(x)$ of X:

Let $0 < x < 1$. Then $f_X(x)=\int_{-\infty}^{\infty} f(x, y)\, dy$

$$=\int_{x-1}^{1-x} \tfrac{1}{2}\, dy = 1-x.$$

Let $-1 < x \leqslant 0$. Then $f_X'(x)=\int_{-\infty}^{\infty} f(x, y)\, dy$

$$=\int_{-x-1}^{x+1} \tfrac{1}{2}\, dy = 1+x.$$

Hence the marginal density function of X is given by

$$f_X'(x)=\begin{cases} 1+x, & \text{for } -1 < x \leqslant 0 \\ 1-x, & \text{for } \quad 0 < x < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Similarly, the marginal density function of Y is given by

$$f_Y'(y)=\begin{cases} 1+y, & \text{for } -1 < y \leqslant 0 \\ 1-y, & \text{for } \quad 0 < y < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Since $f(x, y) \neq f_X'(x)\, f_Y(y)$, X and Y are dependent.

Ex. 8. The joint distribution of X and Y is uniform over the interior of a triangle with vertices $(0, 0)$, $(2, 0)$ and $(1, 2)$. Find the density function of (X, Y) and $P(X \leqslant 1, Y \leqslant 1)$.

Here the probability density function of (X, Y) is given by

$$f(x, y)=\begin{cases} \frac{1}{2}, & \text{if } (x, y) \in R \\ 0, & \text{elsewhere} \end{cases}$$

where R is the interior of the triangle of area 2 square units having vertices $(0, 0)$, $(2, 0)$ and $(1, 2)$.

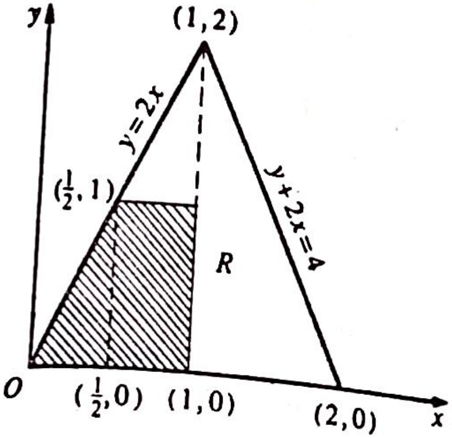


Fig 6.10.3

$\therefore$ $P(X \leqslant 1, Y \leqslant 1)=\int_{-\infty}^{1}\int_{-\infty}^{1} f(x, y)\, dx\, dy$, the region of integration being the shaded portion of Fig. 6.10.3

$$=\int_{0}^{\frac{1}{2}}\left(\int_{0}^{2x} \tfrac{1}{2}\, dy\right) dx+\int_{\frac{1}{2}}^{1}\left(\int_{0}^{1} \tfrac{1}{2}\, dy\right) dx$$

$$=\tfrac{1}{8}+\tfrac{1}{4}=\tfrac{3}{8}.$$

Ex. 9. *A random point (X, Y) is uniformly distributed over a circular region $x^2+y^2<a^2$. Find the marginal distributions of X and Y and the conditional distribution of Y assuming $X=x$, where $|x|<a$.*

The two-dimensional random variable (X, Y) has the density function $f(x, y)$ given by

$$f(x, y)=\begin{cases}\dfrac{1}{\pi a^2}, & x^2+y^2<a^2\\ 0, & \text{elsewhere.}\end{cases}$$

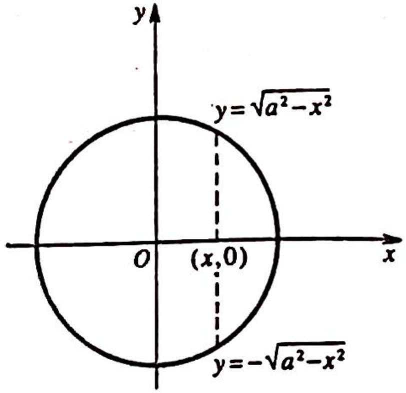


Fig. 6.10.4

The marginal density function $f_X(x)$ of X is given by

$$f_X(x)=\int_{-\infty}^{\infty} f(x, y)\, dy=\int_{-\sqrt{a^2-x^2}}^{\sqrt{a^2-x^2}} \frac{1}{\pi a^2}\, dy$$

$$=\frac{2\sqrt{a^2-x^2}}{\pi a^2},\ |x|<a.$$

$$\therefore\ f_X(x)=\begin{cases}\dfrac{2\sqrt{a^2-x^2}}{\pi a^2}, & -a<x<a\\ 0, & \text{elsewhere.}\end{cases}$$

Similarly the marginal density function $f_Y(y)$ of Y is given by

$$f_Y(y)=\begin{cases}\dfrac{2\sqrt{a^2-y^2}}{\pi a^2}, & -a<y<a\\ 0, & \text{elsewhere.}\end{cases}$$

The conditional density function of Y on the hypothesis $X=x$ is given by

$$f_Y(y/x)=\frac{f(x, y)}{f_X(x)}=\frac{1}{2\sqrt{a^2-x^2}},\ -\sqrt{a^2-x^2}<y<\sqrt{a^2-x^2}.$$

Ex. 10. *The density function of a two-dimensional random variable (X, Y) is given by*

$$f(x, y)=\begin{cases}2, & 0<x<y<1\\ 0, & \text{elsewhere.}\end{cases}$$

Find (i) the marginal density functions of X and Y,

(ii) the conditional density function $f_X(x|y)$ of X, given $Y=y$ and the conditional density function $f_Y(y|x)$ of Y, given $X=x$,

(iii) $P(Y\geqslant \frac{1}{2} \mid X=\frac{1}{2})$ and $P(X\geqslant \frac{1}{3} \mid Y=\frac{2}{3})$.

(i) The marginal density function $f_X(x)$ of X is given by

$$f_X(x)=\int_{-\infty}^{\infty} f(x, y)\, dy=\int_{x}^{1} 2\, dy, \quad \because\ f(x, y)=\begin{cases}2, & \text{for } 0<x<y<1\\ 0, & \text{elsewhere}\end{cases}$$

$$=2(1-x),\ 0<x<1.$$

Similarly, the marginal density function $f_Y(y)$ of Y is given by

$$f_Y(y)=\int_{-\infty}^{\infty} f(x, y)\, dx=\int_{0}^{y} 2dx=2y,\ 0<y<1.$$

Thus, $f_X(x)=\begin{cases}2(1-x), & 0<x<1\\ 0, & \text{elsewhere,}\end{cases}$

and $f_Y(y)=\begin{cases}2y, & 0<y<1\\ 0, & \text{elsewhere.}\end{cases}$

(ii) $f_Y(y|x)=\dfrac{f(x, y)}{f_X(x)}=\dfrac{1}{1-x},\ 0<x<y<1,$

i.e., the conditional distribution of Y on the hypothesis $X=x$ is uniform in $(x\ 1)$.

Again,

$$f_X(x \mid y)=\frac{f(x, y)}{f_Y(y)}=\frac{1}{y},\ 0<x<y<1$$

and so the conditional distribution of X on the hypothesis $Y=y$ is uniform in $(0, y)$.

(iii) $$P(Y > \tfrac{1}{2} \mid X = \tfrac{1}{3}) = \int_{\frac{1}{2}}^{1} f_Y(y \mid \tfrac{1}{3})\, dy = \int_{\frac{1}{2}}^{1} \frac{1}{1-\frac{1}{3}}\, dy = \frac{3}{4}.$$

$$P(X > \tfrac{1}{2} \mid Y = \tfrac{2}{3}) = \int_{\frac{1}{2}}^{\frac{2}{3}} \frac{1}{\frac{2}{3}}\, dx = \frac{1}{4}.$$

Ex. 11. *Let X be uniformly distributed in $(0, 1)$ and let the conditional distribution of Y on the hypothesis $X = x$ be uniform in $(0, x)$. Find the distribution of the two dimensional random variable (X, Y) and the marginal distribution of Y.*

The marginal density function $f_X(x)$ of X is given by

$$f_X(x) = \begin{cases} 1, \text{ for } 0 < x < 1 \\ 0, \quad \text{elsewhere.} \end{cases}$$

Also the conditional density function $f_Y(y \mid x)$ of Y, given $X = x$, is given by

$$f_Y(y \mid x) = \begin{cases} \frac{1}{x}, \text{ for } 0 < y < x < 1 \\ 0, \quad \text{elsewhere.} \end{cases}$$

Since $f(x, y) = f_X(x)\, f_Y(y \mid x)$, we get the density function of (X, Y) as

$$f(x, y) = \begin{cases} \frac{1}{x}, \; 0 < y < x < 1 \\ 0, \quad \text{elsewhere.} \end{cases}$$

Finally, the marginal density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{y}^{1} \frac{1}{x}\, dx = \log \frac{1}{y}, \; 0 < y < 1,$$

i.e., $$f_Y(y) = \begin{cases} -\log y, \; 0 < y < 1 \\ 0 \quad, \text{ elsewhere.} \end{cases}$$

Ex. 12. *Let X, Y be two random variables, each having spectrum $(-\infty, \infty)$. If the conditional density function of X on the hypothesis $Y = y$ is $\frac{|y|\, e^{-x^2 y^2}}{\sqrt{\pi}}$ and the density function of Y is $\frac{\lambda e^{-\lambda^2 y^2}}{\sqrt{\pi}}$, then prove that the density function of X is $\frac{\lambda}{\pi(x^2 + \lambda^2)}$.*

The conditional density function $f_X(x \mid y)$, given $Y = y$, is given by

$$f_X(x \mid y) = \frac{|y|}{\sqrt{\pi}}\, e^{-x^2 y^2}, \; -\infty < y < \infty.$$

and the marginal density function of Y is

$$f_Y(y) = \frac{\lambda}{\sqrt{\pi}}\, e^{-\lambda^2 y^2}, \; -\infty < y < \infty.$$

Since $f(x, y) = f_Y(y)\, f_X(x \mid y)$, we get the density function of (X, Y) as

$$f(x, y) = \frac{\lambda}{\pi}\, |y|\, e^{-y^2(\lambda^2 + x^2)}, \; -\infty < x, y < \infty.$$

$\therefore$ the marginal density function of X is then given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \frac{\lambda}{\pi} \int_{-\infty}^{\infty} |y|\, e^{-y^2(\lambda^2 + x^2)}\, dy$$

$$= \frac{\lambda}{\pi} \left\{ -\int_{-\infty}^{0} y\, e^{-y^2(\lambda^2 + x^2)}\, dy + \int_{0}^{\infty} y\, e^{-y^2(\lambda^2 + x^2)}\, dy \right\}$$

$$= \frac{2\lambda}{\pi} \int_{0}^{\infty} y\, e^{-y^2(x^2 + \lambda^2)}\, dy = \frac{\lambda}{\pi(x^2 + \lambda^2)} \int_{0}^{\infty} e^{-z}\, dz$$

$$= \frac{\lambda}{\pi(x^2 + \lambda^2)}, \; -\infty < x < \infty.$$

Ex. 13. *The joint probability density function of two random variates X, Y is given by*

$$f(x, y) = \begin{cases} \frac{6 - x - y}{8}, \; 0 < x < 2,\, 2 < y < 4 \\ 0 \quad, \text{ elsewhere} \end{cases}$$

Calculate the following probabilities:

(a) $P(X < 1, Y < 3)$, (b) $P(X + Y < 3)$,

(c) $P(X < 1 \mid Y = 3)$, (d) $P(X < 1 \mid Y < 3)$.

The marginal density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \frac{1}{8} \int_{0}^{2} (6 - x - y)\, dx = \frac{5 - y}{4}, \; 2 < y < 4.$$

Hence the conditional density function $f_X(x\,|\,y)$, given $Y=y\ (2<y<4)$, is given by

(a) $$f_X(x\,|\,y)=\frac{f(x,y)}{f_Y(y)}=\frac{\frac{1}{8}(6-x-y)}{\frac{1}{4}(5-y)}=\frac{6-x-y}{2(5-y)},\quad 0<x<2.$$

$$P(X<1,Y<3)=\int_0^1\left\{\int_2^3 f(x,y)\,dy\right\}dx$$

$$=\frac{1}{8}\int_0^1\left\{\int_2^3 (6-x-y)\,dy\right\}dx$$

$$=\frac{1}{16}\int_0^1 (7-2x)\,dx=\frac{3}{8}.$$

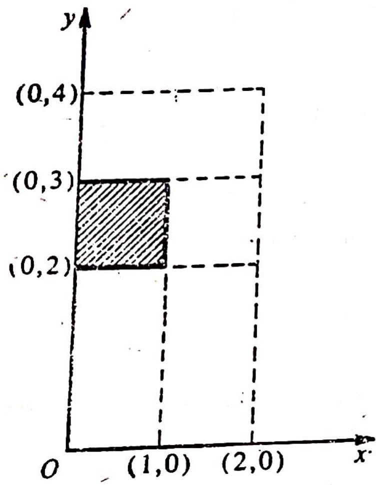


Fig. 6.10.5

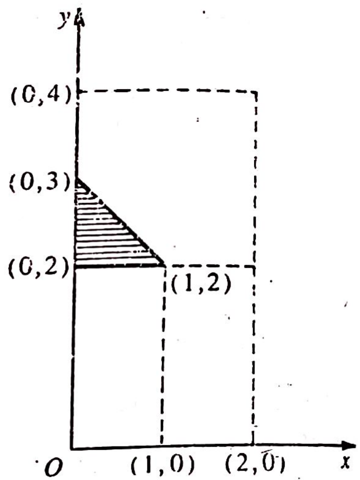


Fig. 6.10.6

(b) $$P(X+Y<3)=P(Y<3-X)$$

$$=\int_0^1\left\{\int_2^{3-x} f(x,y)\,dy\right\}dx$$

$$=\frac{1}{8}\int_0^1\left\{\int_2^{3-x}(6-x-y)\,dy\right\}dx$$

$$=\frac{1}{16}\int_0^1 (x^2-8x+7)\,dx=\frac{5}{24}.$$

(c) $$P(X<1\,|\,Y=3)=\int_0^1 f_X(x\,|\,3)\,dx=\int_0^1\frac{f(x,3)}{f_Y(3)}\,dx$$

$$=\int_0^1\frac{\frac{6-x-3}{8}}{\frac{5-3}{4}}\,dx=\frac{1}{4}\int_0^1 (3-x)\,dx=\frac{5}{8}.$$

(d) $$P(X<1\,|\,Y<3)=\frac{P(X<1,Y<3)}{P(Y<3)}.$$

Now, $$P(Y<3)=\int_2^3\frac{5-y}{4}\,dy=\frac{5}{8}$$

and so $$P(X<1\,|\,Y<3)=\frac{\frac{3}{8}}{\frac{5}{8}}=\frac{3}{5}.$$

Ex. 14. *Suppose (X, Y) is uniformly distributed over the area bounded by $y^2=x$ and $x=4$. Find the joint distribution of X and Y and $P(X<3,\ Y<0)$.*

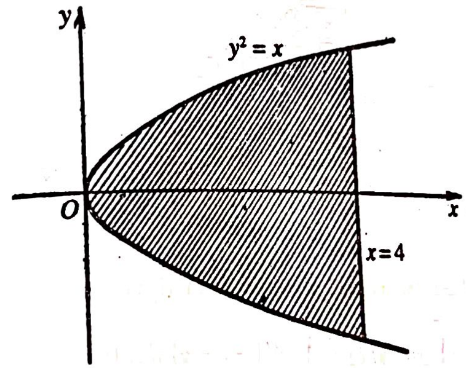


Fig. 6.10.7

Here the density function of (X, Y) can be taken as

$$f(x,y)=\begin{cases} C, & (x,y)\in R\\ 0, & \text{elsewhere}\end{cases}$$

where R is the shaded region shown in Fig. 6.10.7 and C is a constant. Then from $\iint_R f(x,y)\,dx\,dy=1$, we get

$$\int_{-2}^{2}\left(\int_{y^2}^{4} C\,dx\right)dy=1 \quad\text{or,}\quad C\int_{-2}^{2}(4-y^2)\,dy=1.\quad \therefore\ C=\tfrac{3}{32}.$$

Now, $P(X<3, Y<0)=\int_0^3 \left\{ \int_{-\sqrt{x}}^0 f(x, y)\, dy \right\} dx$

$$=\frac{3}{32}\int_0^3 \sqrt{x}\, dx=\frac{3\sqrt{3}}{16}.$$

Ex. 15. *The random variables X and Y have the joint density function*

$$f(x, y)=\begin{cases} 6(1-x-y), & \text{for } x>0, y>0, x+y<1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the marginal distributions of X and Y. Are X and Y independent? [C. H. (Math.) '83, '86]

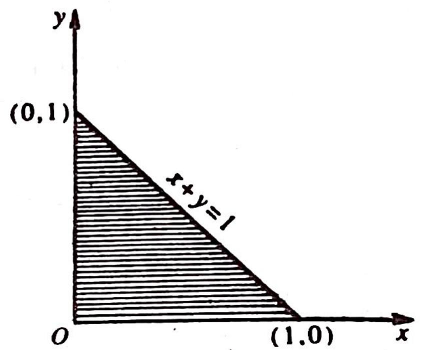


Fig. 6.10.8

The marginal density function of X is given by

$$f_X(x)=\int_{-\infty}^{\infty} f(x, y)\, dy=\int_0^{1-x} 6(1-x-y)\, dy,\ 0<x<1$$

$$=3(1-x)^2,\ 0<x<1.$$

Similarly, the marginal density function of Y is given by

$$f_Y(y)=3(1-y)^2,\ 0<y<1.$$

Since $f(x, y) \neq f_X(x) f_Y'(y)$, X and Y are dependent.

Ex. 16. *The density function of the two-dimensional random variable (X, Y) is given by*

$$f(x, y)=\begin{cases} C(2x+5y), & 0\leq x\leq 3,\ 2\leq y\leq 4 \\ 0 & ,\ \text{elsewhere.} \end{cases}$$

Find (a) the value of C, (b) the marginal dedsity functions of X and Y, (c) the joint distribution function F(x, y), (d) P(X+Y < 3), (e) $f_X'(x \mid y)$. Are X and Y independent?

(a) From $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy=1$, we get

$$C\int_0^3 \left\{ \int_2^4 (2x+5y)\, dy \right\} dx=1.$$

$$\therefore\ C=\frac{1}{108}.$$

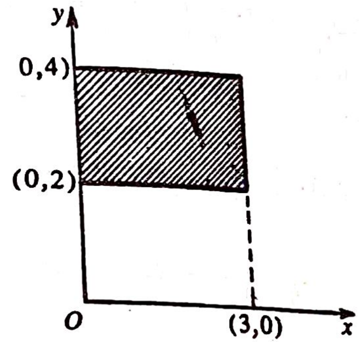


Fig. 6.10.9

(b) The marginal density function of X is given by

$$f_X'(x)=\frac{1}{108}\int_{-\infty}^{\infty} f(x, y)\, dy$$

$$=\frac{1}{108}\int_2^4 (2x+5y)\, dy$$

$$=\frac{1}{108}(4x+30),\ 0\leq x\leq 3.$$

The marginal density function of Y is given by

$$f_Y'(y)=\frac{1}{108}\int_{-\infty}^{\infty} f(x, y)\, dx=\frac{1}{108}\int_0^3 (2x+5y)\, dx$$

$$=\frac{1}{108}(9+15y),\ 2\leq y\leq 4.$$

(c) The joint distribution function $F(x, y)$ is given by

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x', y')\, dx'\, dy'$$

$$= \int_{0}^{x}\left\{\int_{2}^{y} \frac{1}{108}\,(2x' + 5y')\, dy'\right\} dx'$$

$$= \frac{1}{216}\, x(y-2)(2x+5y+10) \quad \text{if } 0 \leqslant x \leqslant 3,\ 2 \leqslant y \leqslant 4.$$

$$= \tfrac{1}{72}\,(y-2)\,(5y+16) \quad \text{if } x > 3,\ 2 \leqslant y \leqslant 4$$

$$= \tfrac{1}{54}\, x(x+15) \quad \text{if } 0 \leqslant x \leqslant 3,\ y > 4$$

$$= 0 \quad \text{if } x < 0 \text{ or } y < 2$$

$$= 1 \quad \text{if } x > 3,\ y > 4.$$

(d) $$P(X+Y<3) = \int_{0}^{1}\left\{\int_{2}^{3-x} \frac{1}{108}\,(2x+5y)\, dy\right\} dx$$

$$= \frac{1}{216}\int_{0}^{1} (x^2 - 26x + 25)\, dx$$

$$= \frac{37}{648}.$$

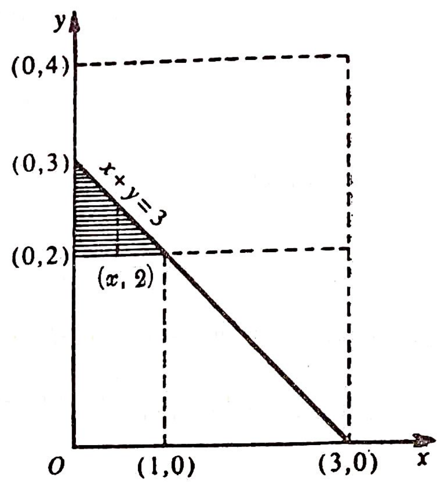


Fig. 6.10.10

(e) $$f_X(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

$$= \frac{\frac{1}{108}\,(2x+5y)}{\frac{1}{108}\,(9+15y)} = \frac{2x+5y}{9+15y},\ 0 < x < 3,\ 2 < y < 4.$$

Here $f(x, y) \neq f_X(x)\, f_Y(y)$. So X, Y are dependent.

Ex. 17. *The joint density function of X and Y is given by*

$$f(x, y) = \begin{cases} k(x+y), & \text{for } 0 < x < 1,\ 0 < y < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find (a) the value of k, (b) the marginal density function $f_X(x)$ and $f_Y(y)$, (c) the conditional density functions $f_Y(y \mid x)$, $f_X(x \mid y)$, (d) the conditional distribution function $F_Y(y \mid x)$, (e) $P(\mid X - Y \mid < \frac{1}{2})$. Are X and Y independent?

(a) From $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$, we get

$$\int_{0}^{1}\int_{0}^{1} k(x+y)\, dx\, dy = 1$$

or, $$k\int_{0}^{1} (\tfrac{1}{2} + y)\, dy = 1.$$

$\therefore\ k = 1.$

(b) The marginal density function $f_X(x)$ is given by,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \int_{0}^{1} (x+y)\, dy = x + \tfrac{1}{2},\ 0 < x < 1.$$

Similarly, $f_Y(y) = y + \frac{1}{2},\ 0 < y < 1.$

(c) The conditional density functions are

$$f_Y(y \mid x) = \frac{f(x, y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}},\ 0 < y < 1,$$

for any fixed x $(0 < x < 1)$ and

$$f_X(x \mid y) = \frac{f(x, y)}{f_Y(y)} = \frac{x+y}{y+\frac{1}{2}},\ 0 < x < 1,$$

for any fixed y $(0 < y < 1)$.

(d) The conditional distribution function $F_Y(y \mid x)$ is given by

$$F_Y(y \mid x) = \int_{-\infty}^{y} f_Y(t \mid x)\, dt$$

$$= \int_0^y \frac{x+t}{x+\frac{1}{2}}\, dt = \frac{xy + \frac{y^2}{2}}{x+\frac{1}{2}}, \text{ if } 0 < y < 1$$

$= 1$, if $y \geqslant 1$

$= 0$, elsewhere.

(e) We first find $P(|X-Y| > \frac{1}{2})$.

Now, $P(|X-Y| > \frac{1}{2}) = P(X > Y, X-Y > \frac{1}{2}) + P(X \leqslant Y, Y-X > \frac{1}{2})$

$(\because |X-Y| > \frac{1}{2}$ implies, $X > Y, X-Y > \frac{1}{2}$, or, $X < Y, Y-X > \frac{1}{2})$

$= P\{(X, Y) \in R_1 \cup R_2\}$,

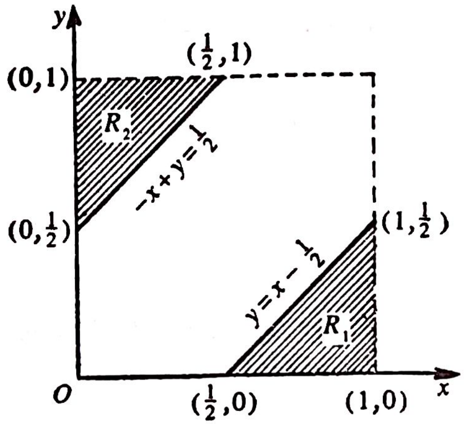


Fig. 6.10.11

where R_1 and R_2 are the two shaded regions, shown in Fig. 6.10.11

$$= \int_{\frac{1}{2}}^{1} \left\{ \int_0^{x-\frac{1}{2}} (x+y)\, dy \right\} dx + \int_0^{\frac{1}{2}} \left\{ \int_{x+\frac{1}{2}}^{1} (x+y)\, dy \right\} dx$$

$$= \int_{\frac{1}{2}}^{1} \left(\frac{3x^2}{2} - x + \frac{1}{8} \right) dx + \int_0^{\frac{1}{2}} \left(-\frac{3}{2}x^2 + \frac{3}{8} \right) dx = \frac{1}{8} + \frac{1}{8} = \frac{1}{4},$$

$\therefore \quad P(|X-Y| \leqslant \frac{1}{2}) = \frac{3}{4}$.

Since $f(x, y) \neq f_X(x) f_Y(y)$, X and Y are dependent.

Ex. 18. *The probability density function of a two-dimensional random variable (X, Y) is given by*

$$f(x, y) = \begin{cases} k(x+y), & x > 0, y > 0, x+y < 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find k and $P(X < 1, Y > \frac{1}{2})$.

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$ gives $\int_0^2 \left\{ \int_0^{2-x} k(x+y)\, dy \right\} dx = 1$

or, $\frac{k}{2} \int_0^2 (4 - x^2)\, dx = 1. \quad \therefore \quad k = \frac{3}{8}$.

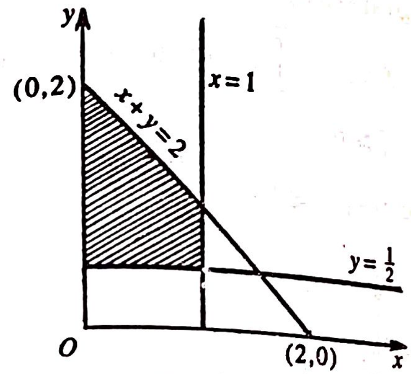


Fig. 6.10.12

Now, $P(X < 1, Y > \frac{1}{2})$

$= P\{(X, Y) \in R\}$ where R is the shaded region of Fig. 6.10.12

$$= \frac{3}{8} \int_0^1 \left\{ \int_{\frac{1}{2}}^{2-x} (x+y)\, dy \right\} dx = \frac{3}{64} \int_0^1 (15 - 4x - 4x^2)\, dx = \frac{35}{64}.$$

Ex. 19. *The distribution of a two-dimensional random variable (X, Y) is given by*

$$f(x, y) = \begin{cases} e^{-x-y}, & x \geqslant 0, y \geqslant 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find (a) the joint distribution function, (b) the marginal distribution functions of X and Y, (c) $P(X = Y)$, (d) $P(X+Y \leqslant 4)$, (e) $P(X \geqslant 1)$, (f) $P(X \leqslant Y)$, (g) $P(a < X+Y < b)$ where $0 < a < b$. Show also that X and Y are independent.

(a) The joint distribution function $F(x, y)$ is given by

$$F(x, y) = \int_{-\infty}^{y} \left\{ \int_{-\infty}^{x} f(x', y')\, dx' \right\} dy'$$

$$= \left(\int_{0}^{x} e^{-x'}\, dx' \right) \left(\int_{0}^{y} e^{-y'}\, dy' \right)$$

$$= (1-e^{-x})(1-e^{-y}), \text{ if } x \geqslant 0,\ y \geqslant 0.$$

$$\therefore\ F(x, y) = \begin{cases} (1-e^{-x})(1-e^{-y}), & \text{if } x \geqslant 0,\ y \geqslant 0 \\ 0, & \text{elsewhere.} \end{cases}$$

(b) The marginal density functions $f_X(x)$ and $f_Y(y)$ are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{0}^{\infty} e^{-x} \cdot e^{-y}\, dy = e^{-x},\ x \geqslant 0$$

and $f_Y(y) = e^{-y},\ y \geqslant 0.$

(c) $P(X+Y \leqslant 4) = P(Y \leqslant 4-X) = \int_{0}^{4} e^{-x} \left\{ \int_{0}^{4-x} e^{-y}\, dy \right\} dx$

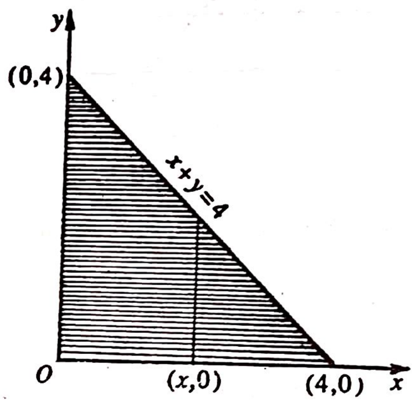


Fig. 6.10.13

$$= \int_{0}^{4} e^{-x}\left(1-e^{-4+x}\right) dx = \int_{0}^{4} e^{-x}\, dx - \int_{0}^{4} e^{-4}\, dx$$

$$= 1-e^{-4} - 4e^{-4} = 1-5e^{-4}.$$

(d) $P(X \geqslant 1) = \int_{1}^{\infty} f_X(x)\, dx = \int_{1}^{\infty} e^{-x}\, dx = e^{-1}.$

(e) $P(X \leqslant Y) = \int_{0}^{\infty} e^{-x} \left(\int_{x}^{\infty} e^{-y}\, dy \right) dx = \int_{0}^{\infty} e^{-2x}\, dx = \frac{1}{2}.$

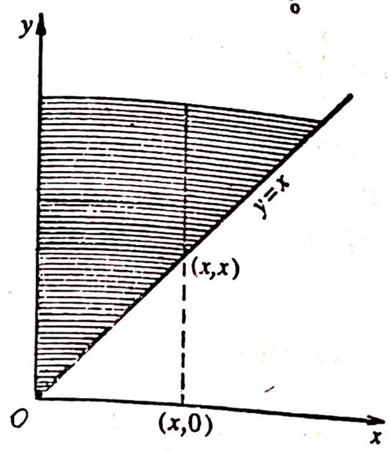


Fig. 6.10.14

(f) $P(a \leqslant X+Y \leqslant b),\ 0 < a < b$

$= P(a-X \leqslant Y \leqslant b-X) = P\{(X, Y) \in R_1 \cup R_2\},$

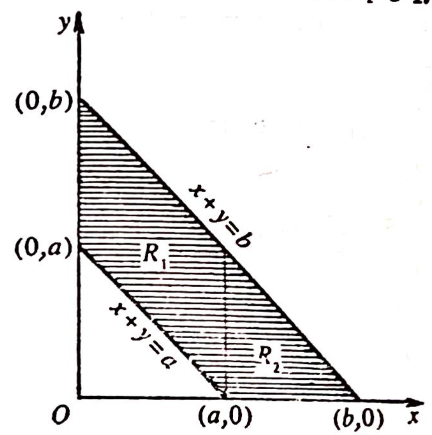


Fig. 6.10.15

where R_1 and R_2 are the shaded regions shown in Fig. 6.10.15

$$= \int_{0}^{a} e^{-x} \left(\int_{a-x}^{b-x} e^{-y}\, dy \right) dx + \int_{a}^{b} e^{-x} \left(\int_{0}^{b-x} e^{-y}\, dy \right) dx$$

$$=\int_0^a e^{-x}\left(e^{-a+x}-e^{-b+x}\right)dx+\int_a^b e^{-x}\left(1-e^{-b+x}\right)dx$$

$$=a(e^{-a}-e^{-b})+\int_a^b (e^{-x}-e^{-b})\,dx$$

$$=a(e^{-a}-e^{-b})-(e^{-b}+be^{-b}-e^{-a}-ae^{-b})$$

$$=e^{-a}(a+1)-e^{-b}(b+1).$$

Since $f(x, y)=f_X(x)\,f_Y(y)$ for all x, y, the random variables X and Y are independent.

Ex. 20. *The density function of a two-dimensional random variable (X, Y) is given by*

$$f(x, y)=\begin{cases} kxy(x+y), & 0<x<1,\ 0<y<1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find (a) the value of k, (b) the marginal density function of X and Y, (c) $P(\frac{1}{2}<X<\frac{3}{4}, \frac{1}{3}<Y<\frac{2}{3})$, (d) $f_X(x\mid y), f_Y(y\mid x)$.

(a) From $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x, y)\,dx\,dy=1$, we get

$$\int_0^1\int_0^1 kxy(x+y)\,dx\,dy=1$$

or, $k\int_0^1\left(\frac{x^2}{2}+\frac{x}{3}\right)dx=1$;

$\therefore\ k=3.$

(b) $$f_X(x)=\int_{-\infty}^{\infty} f(x, y)\,dy$$

$$=3\int_0^1 xy(x+y)\,dy$$

$$=\tfrac{3}{2}x^2+x,\ 0<x<1.$$

Similarly, $f_Y(y)=\frac{3}{2}y^2+y,\ 0<y<1.$

(c) $P(\frac{1}{2}<X<\frac{3}{4}, \frac{1}{3}<Y<\frac{2}{3})$

$$=3\int_{1/2}^{3/4}\left\{\int_{1/3}^{2/3} xy(x+y)\,dy\right\}dx$$

$$=3\int_{1/2}^{3/4}\left(\frac{x^2}{6}+\frac{7x}{81}\right)dx=\frac{311}{3456}.$$

(d) $$f_Y(y\mid x)=\frac{f(x, y)}{f_X(x)}=\frac{3xy(x+y)}{\frac{3}{2}x^2+x}=\frac{6y(x+y)}{2+3x},\ 0<y<1.$$

Similarly, $f_X(x\mid y)=\dfrac{6x(x+y)}{3y+2},\ 0<x<1.$

Ex. 21. *The density function of a two-dimensional random variable (X, Y) is given by*

$$f(x, y)=k(1-x^2-y^2),\ 0<x^2+y^2<1.$$

Find k and the marginal density functions of X and Y.

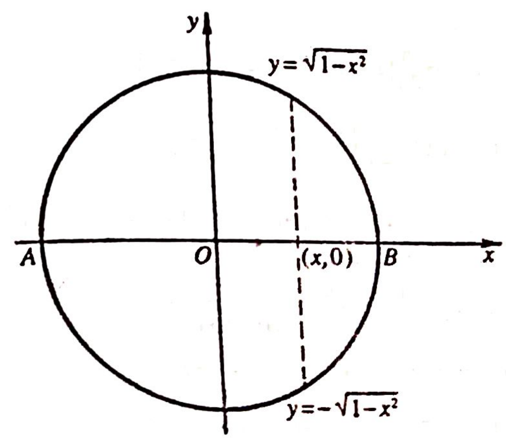


Fig. 6.10.16

From $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(x, y)\,dx\,dy=1$, we get

$k\iint\limits_R (1-x^2-y^2)\,dx\,dy=1$, where R is the circular region $\{(x, y): x^2+y^2<1\}$ (Fig. 6.10.16).

Transforming to polar co-ordinates,

$x = r\cos\theta,\ y = r\sin\theta,$ so that $\left|\dfrac{\partial(x,y)}{\partial(r,\theta)}\right| = r$, we get

$$k\int_0^1\int_0^{2\pi}(1-r^2)\,r\,dr\,d\theta = 1$$

or, $2\pi k\left(\frac{1}{2}-\frac{1}{4}\right) = 1. \quad \therefore\ k = \frac{2}{\pi}.$

Now, $f_X(x) = \dfrac{2}{\pi}\displaystyle\int_{-\infty}^{\infty} f(x, y)\,dy$

$$= \frac{2}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}(1-x^2-y^2)\,dy$$

$$= \frac{2}{\pi}\left\{2(1-x^2)^{\frac{3}{2}} - \frac{2}{3}(1-x^2)^{\frac{3}{2}}\right\}$$

$$= \frac{8}{3\pi}\left(1-x^2\right)^{\frac{3}{2}},\quad -1 < x < 1.$$

Similarly, $f_Y(y) = \dfrac{8}{3\pi}\left(1-y^2\right)^{\frac{3}{2}},\ -1 < y < 1.$

Ex. 22. *The density function of the two-dimensional random variable (X, Y) is given by*

$$f(x, y) = \begin{cases} 3x^2y + 3y^2x, & 0 < x < 1,\ 0 < y < 1 \\ 0, & \textit{elsewhere.}\end{cases}$$

Find the marginal density functions and $P(\frac{1}{3} < Y \leq \frac{2}{3} \mid \frac{1}{2} < X \leq \frac{3}{4})$.

The marginal density function $f_X(x)$ is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy = \int_0^1 (3x^2y + 3y^2x)\,dy$$

$$= \frac{3x^2}{2} + x,\ \text{for } 0 < x < 1.$$

Similarly, $f_Y(y) = \dfrac{3y^2}{2} + y$, for $0 < y < 1$.

Then $P(\frac{1}{3} < Y \leq \frac{2}{3} \mid \frac{1}{2} < X \leq \frac{3}{4})$

$$= \frac{P(\frac{1}{2} < X \leq \frac{3}{4},\ \frac{1}{3} < Y \leq \frac{2}{3})}{P(\frac{1}{2} < X \leq \frac{3}{4})}$$

$$= \frac{\displaystyle\int_{\frac{1}{3}}^{\frac{2}{3}}\left\{\int_{\frac{1}{2}}^{\frac{3}{4}} f(x, y)\,dx\right\}dy}{\displaystyle\int_{\frac{1}{2}}^{\frac{3}{4}} f_X(x)\,dx} = \frac{\displaystyle\int_{\frac{1}{3}}^{\frac{2}{3}}\left\{\int_{\frac{1}{2}}^{\frac{3}{4}} (3x^2y + 3y^2x)\,dx\right\}dy}{\displaystyle\int_{\frac{1}{2}}^{\frac{3}{4}}\left(\frac{3x^2}{2} + x\right)dx}$$

$$= \frac{\displaystyle\int_{\frac{1}{3}}^{\frac{2}{3}}\left(\frac{19y}{64} + \frac{15y^2}{32}\right)dy}{\frac{39}{128}} = \frac{311}{1053}.$$

Ex. 23. *The density function of a two-dimensional random variable (X, Y) is $f(x, y) = C\,e^{-\frac{x^2-xy+y^2}{2}},\ -\infty < x, y < \infty$. Find C and show that the marginal distributions of X and Y are both normal $\left(0, \frac{2}{\sqrt{3}}\right)$. Show also that the conditional distribution of Y given $X = x$ is normal $\left(\frac{x}{2}, 1\right)$.*

From $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$, we get

$$C\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} e^{-\frac{x^2-xy+y^2}{2}}\,dy\right)dx = 1$$

$$\text{or,}\quad C\int_{-\infty}^{\infty} e^{-\frac{3x^2}{8}}\left\{\int_{-\infty}^{\infty} e^{-\frac{\left(y-\frac{x}{2}\right)^2}{2}}\,dy\right\}dx = 1. \qquad (6.10.1)$$

We now put $z = y - \frac{x}{2}$. Then

$$\int_P^Q e^{-\frac{\left(y-\frac{x}{2}\right)^2}{2}}\,dy = \int_{P-\frac{x}{2}}^{Q-\frac{x}{2}} e^{-\frac{z^2}{2}}\,dz,$$

so that, proceeding to the limits $P \to -\infty,\ Q \to \infty$.

$$\int_{-\infty}^{\infty} e^{-\frac{\left(y-\frac{x}{2}\right)^2}{2}}\, dy = \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\, dz = \sqrt{2\pi}. \qquad (6.10.2)$$

$\therefore$ From (6. 10. 1), we get

$$C\sqrt{2\pi}\int_{-\infty}^{\infty} e^{-\frac{3x^2}{8}}\, dx = 1. \qquad (6.10.3)$$

Now, again putting $\frac{\sqrt{3}x}{2} = t$, we get

$$\int_{P}^{Q} e^{-\frac{3x^2}{8}}\, dx = \frac{2}{\sqrt{3}}\int_{\frac{\sqrt{3}P}{2}}^{\frac{\sqrt{3}Q}{2}} e^{-\frac{t^2}{2}}\, dt.$$

Hence proceeding to the limits $P \to -\infty$, $Q \to \infty$, we get

$$\int_{-\infty}^{\infty} e^{-\frac{3x^2}{8}}\, dx = \frac{2}{\sqrt{3}}\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt = \frac{2}{\sqrt{3}}\sqrt{2\pi}.$$

$\therefore$ From (6. 10.3), we get

$$C \cdot 2\pi \cdot \frac{2}{\sqrt{3}} = 1, \text{ i.e., } C = \frac{\sqrt{3}}{4\pi}.$$

Now the marginal density function of X is

$$f_X(x) = \frac{\sqrt{3}}{4\pi}\int_{-\infty}^{\infty} e^{-\frac{(x^2-xy+y^2)}{2}}\, dy$$

$$= \frac{\sqrt{3}}{4\pi} e^{-\frac{3x^2}{8}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(y-\frac{x}{2}\right)^2}\, dy$$

$$= \frac{\sqrt{3}}{4\pi}\sqrt{2\pi}\, e^{-\frac{3x^2}{8}}, \text{ by } (6.10.2)$$

$$= \frac{1}{\sqrt{2\pi}\cdot\frac{2}{\sqrt{3}}} e^{-\frac{x^2}{2\left(\frac{2}{\sqrt{3}}\right)^2}}, \quad -\infty < x < \infty.$$

Similarly, $f_Y(y) = \dfrac{1}{\sqrt{2\pi}\cdot\frac{2}{\sqrt{3}}} e^{-\frac{y^2}{2\left(\frac{2}{\sqrt{3}}\right)^2}}, \quad -\infty < y < \infty.$

Thus both X and Y are normally distributed with parameters $\left(0, \frac{2}{\sqrt{3}}\right)$.

Again the conditional density function $f_Y(y \mid x)$ of Y, given $X = x$, is given by

$$f_Y(y \mid x) = \frac{f(x, y)}{f_X(x)} = \frac{\frac{\sqrt{3}}{4\pi} e^{-\frac{(x^2-xy+y^2)}{2}}}{\frac{\sqrt{3}}{2\sqrt{2\pi}} e^{-\frac{3x^2}{8}}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(y-\frac{x}{2}\right)^2}{2}}, \quad \infty < y < \infty,$$

which shows that the conditional distribution of Y, given $X = x$, is normal $\left(\frac{x}{2}, 1\right)$.

Ex. 24. *Let* $f(x, y) = \begin{cases} C(y-x)^{\alpha}, & 0 \leqslant x < y < 1 \\ 0 & , \textit{elsewhere.} \end{cases}$

(a) *What values of* α *and* C *should be chosen so that* f *is a possible density function ?*

(b) *Find the marginal density functions.*

From $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$, we get

$$\int_{0}^{1}\left\{\int_{x}^{1} C(y-x)^{\alpha}\, dy\right\} dx = 1$$

or, $$\frac{C}{\alpha+1}\int_{0}^{1}(1-x)^{\alpha+1}\, dx = 1.$$

$\therefore \quad C = (\alpha+1)(\alpha+2).$

$\therefore \quad f(x, y) = \dfrac{1}{(\alpha+1)(\alpha+2)}(y-x)^{\alpha}, \; 0 \leqslant x < y < 1.$

Now $f(x, y) \geqslant 0$ for all x, y implies $\alpha+1 > 0$, i.e., $\alpha > -1$ or, $\alpha < -2$.

Hence, $f(x, y)$ to be a possible density function, we must have

$C=(\alpha+1)(\alpha+2), \alpha>-1$ or, $\alpha<-2$

(b) $f_X'(x)=\int_{-\infty}^{\infty} f(x, y)\, dy$

$$=(\alpha+1)(\alpha+2)\int_x^1 (y-x)^{\alpha}\, dy$$

$$=(\alpha+2)(1-x)^{\alpha+1}.$$

$$\therefore\ f_X(x)=\begin{cases}(\alpha+2)(1-x)^{\alpha+1}, & 0<x<1\\ 0, & \text{elsewhere.}\end{cases}$$

Similarly, $f_Y(y)=\begin{cases}(\alpha+2)y^{\alpha+1}, & 0<y<1\\ 0, & \text{elsewhere.}\end{cases}$

Ex. 25. *If independent random variables X and Y be each uniformly distributed in the interval $(-a, a)$, then find the distribution of (a) $X+Y$, (b) XY and (c) $\frac{X}{Y}$.*

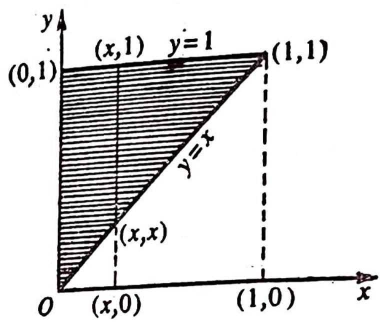


Fig. 6.10.17

(a) X and Y being each uniformly distributed in $(-a, a)$ their density functions $f_X(x)$ and $f_Y(y)$ are respectively given by

$$f_X(x)=\begin{cases}\frac{1}{2a}, & -a<x<a\\ 0, & \text{elsewhere}\end{cases}$$

and $f_Y(y)=\begin{cases}\frac{1}{2a}, & -a<y<a\\ 0, & \text{elsewhere.}\end{cases}$

Since X and Y are independent, the density function of (X, Y) is given by $f_{X,Y}(x, y)=\begin{cases}\frac{1}{4a^2}, & -a<x<a,\ -a<y<a\\ 0, & \text{elsewhere.}\end{cases}$

We put $U=X+Y$, $V=X$.

In terms of real variables, $u=x+y$, $v=x$.

$$\therefore\ \frac{\partial(u, v)}{\partial(x, y)}=\begin{vmatrix}1 & 1\\ 1 & 0\end{vmatrix}=-1<0 \text{ for all } x, y.$$

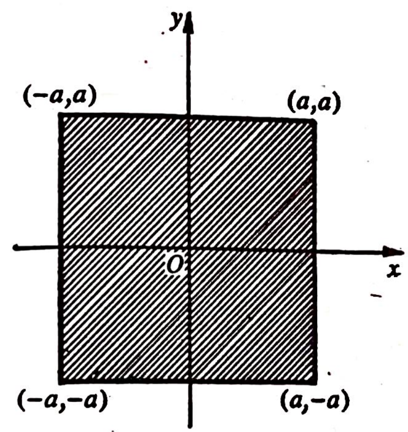


Fig. 6.10.18

Hence, if $f_{V,U}(u, v)$ be the joint density function of the random variables U and V, then

$$f_{U,V}(u, v)=f_{X,Y}(x, y)\cdot 1$$

$$=f_X(x)\, f_Y(y),\ X \text{ and } Y \text{ being independent.}$$

$$=\frac{1}{4a^2}, \text{ if } -a<x<a,\ -a<y<a$$

$$i.e., \text{ if } -a<v<a,\ -a<u-v<a.$$

It is then evident that as the point (x, y) varies within the square region (in the xy plane) having vertices $(-a, -a)$, $(a, -a)$, (a, a) and $(-a, a)$, the point (u, v) varies within the region in the

$u\,v$ plane bounded by the lines $v=-a$, $u-v=a$, $v=a$ and $v-u=a$. Let $f_U(u)$ be the probability density function of the random variable U.

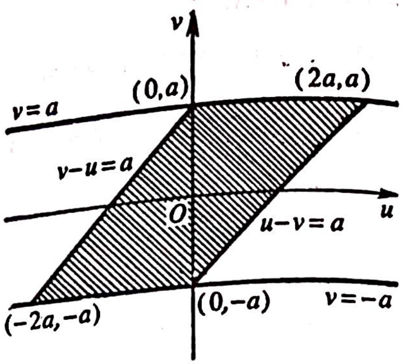


Fig. 6.10.19

Let $-2a < u < 0$.

Then $f_U(u)=\displaystyle\int_{-a}^{u+a}\frac{1}{4a^2}\,dv=\frac{u+2a}{4a^2}$.

Let $0 \leqslant u < 2a$

Then $f_U(u)=\displaystyle\int_{u-a}^{a}\frac{1}{4a^2}\,dv=\frac{2a-u}{4a^2}$.

Thus the density function $f_U(u)$ is given by

$$f_U(u)=\begin{cases}\dfrac{u+2a}{4a^2}, & -2a<u<0\\[2ex] \dfrac{2a-u}{4a^2}, & 0\leqslant u<2a\end{cases}$$

(*b*) We put $W=XY$ and $V=X$.

In terms of real variables, $w=xy$, $v=x$.

Then $\dfrac{\partial(w,v)}{\partial(x,y)}=\begin{vmatrix} y & x\\ 1 & 0\end{vmatrix}=-x$, which changes sign as x varies in the interval $(-a,a)$. We, therefore, proceed in the following way.

If $F_W(w)$ be the distribution function of W, then

$$F_W(w)=P(-\infty<W\leqslant w)=P(-\infty<XY\leqslant w).$$

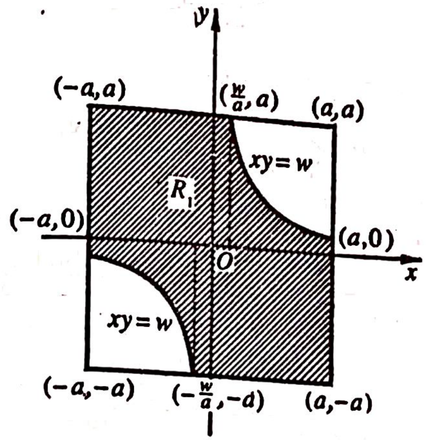


Fig. 6.10.20

Let $w>0$. Then $F_W(w)=\displaystyle\iint_{R_1} f(x,y)\,dx\,dy$,

where R_1 is the shaded region shown in Fig. 6.10.20.

$$=\int_{-a}^{-\frac{w}{a}}\left(\int_{\frac{w}{x}}^{a}\frac{1}{4a^2}\,dy\right)dx+\int_{-\frac{w}{a}}^{\frac{w}{a}}\left(\int_{-a}^{a}\frac{1}{4a^2}\,dy\right)dx+\int_{\frac{w}{a}}^{a}\left(\int_{-a}^{\frac{w}{x}}\frac{1}{4a^2}\,dy\right)dx$$

$$=\frac{1}{4a^2}\left\{\int_{-a}^{-\frac{w}{a}}\left(a-\frac{w}{x}\right)dx+\int_{-\frac{w}{a}}^{\frac{w}{a}}2a\,dx+\int_{\frac{w}{a}}^{a}\left(\frac{w}{x}+a\right)dx\right\}$$

$$=\frac{1}{4a^2}\left\{\left(-w-w\log\frac{w}{a^2}+a^2\right)+4w+\left(a^2-w\log\frac{w}{a^2}-w\right)\right\}$$

$$=\frac{1}{4a^2}\left(2w+2a^2-2w\log\frac{w}{a^2}\right).$$

Hence, the density function of the random variable W is

$$f_W(w)=F'_W(w)=\frac{1}{2a^2}\left(1-\log\frac{w}{a^2}-w.\frac{1}{w}\right)$$

$$=\frac{1}{2a^2}\log\frac{a^2}{w}.\ w>0.$$

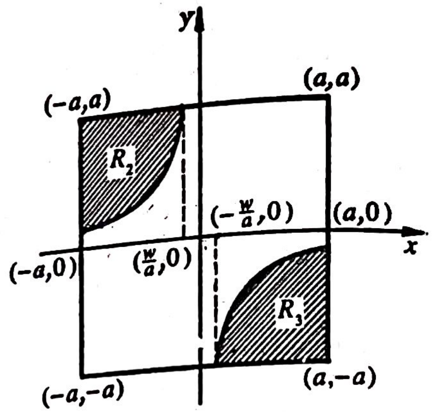


Fig. 6.10.21

Let $w<0$. In this case,

$$F_W(w)=\iint_{R_2\cup R_3} f(x,y)\,dx\,dy$$

where R_2 and R_3 is the shaded region in Fig. 6.10.21.

$$\therefore\quad F_W(w)=\int_{-a}^{\frac{w}{a}}\left(\int_{\frac{w}{x}}^{a}\frac{1}{4a^2}dy\right)dx+\int_{-\frac{w}{a}}^{a}\left(\int_{-a}^{\frac{w}{x}}\frac{1}{4a^2}dy\right)dx$$

$$=\frac{1}{4a^2}\left\{\int_{-a}^{\frac{w}{a}}\left(a-\frac{w}{x}\right)dx+\int_{-\frac{w}{a}}^{a}\left(a+\frac{w}{x}\right)dx\right\}$$

$$=\frac{1}{4a^2}\left[\left\{w+a^2-w\log\left(-\frac{w}{a^2}\right)\right\}+\left\{a^2+w-w\log\left(-\frac{w}{a^2}\right)\right\}\right]$$

$$=\frac{1}{2a^2}\left[a^2+w-w\log\left(-\frac{w}{a^2}\right)\right].$$

Hence as before the density function, in this case is given by,

$$f_W(w)=\frac{1}{2a^2}\log\left(-\frac{a^2}{w}\right),\ w<0.$$

Hence the p. d. f. of W is

$$f_W(w)=\begin{cases}\dfrac{1}{2a^2}\log\left(\dfrac{a^2}{w}\right), & 0<w<a^2\\[2mm] \dfrac{1}{2a^2}\log\left(-\dfrac{a^2}{w}\right), & -a^2<w<0.\end{cases}$$

(c) Let $Z=\dfrac{X}{Y}$.

In terms of real variables, $z=\dfrac{x}{y}$. If $F_Z(z)$ be the distribution function of Z, then

$$F_Z(z)=P\left(-\infty<\frac{X}{Y}<z\right).$$

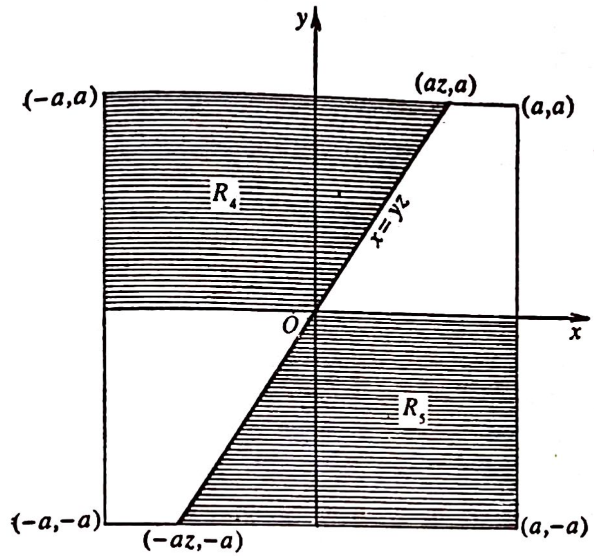


Fig. 6.10.22

Let $0<z<1$. Then $F_Z(z)=\iint_{R_4\cup R_5} f(x,y)\,dx\,dy$

where R_4 and R_5 are shaded regions shown in Fig. 6.10.22.

$$\therefore\quad F_z(z)=\int_0^{az}\left\{\int_{\frac{x}{z}}^{a} f(x,y)\,dy\right\}dx+\int_{-a}^{0}\left\{\int_0^a f(x,y)\,dy\right\}dx$$

$$+\int_{-az}^{0}\left\{\int_{-a}^{\frac{x}{z}} f(x,y)\,dy\right\}dx+\int_0^a\left\{\int_{-a}^{0} f(x,y)\,dy\right\}dx,$$

where the four integrals are successively calculated over the parts of the regions of integration lying in the first, second, third and fourth quadrants respectively.

$$\therefore\quad F_z(z)=\frac{1}{4a^2}\left\{\int_0^{az}\left(a-\frac{x}{z}\right)dx+\int_{-a}^{0}a\,dx+\int_{-az}^{0}\left(\frac{x}{z}+a\right)dx+\int_0^a a\,dx\right\}$$

$$=\frac{1}{4a^2}\left(\frac{a^2z}{2}+a^2+\frac{a^2z}{2}+a^2\right)=\frac{1}{4}(z+2),\ 0<z<1$$

Hence we get

$$F_z'(z)=\tfrac{1}{4},\ 0<z<1$$

Let $z \geqslant 1$.

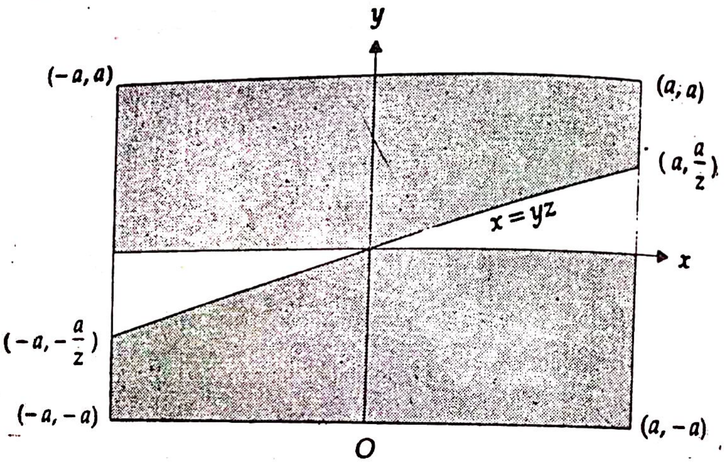


Fig. 6.10.23

As in the case $0 \leqslant z < 1$. we similarly find [Fig. 6.10.23] that

$$F_z(z)=\frac{1}{4a^2}\left[2a^2-\frac{a^2}{2z}+2a^2-\frac{a^2}{2z}\right]$$

that is, $F_z(z)=1-\dfrac{1}{4z}$ if $z \geqslant 1$,

Hence we get $F_z'(z)=\dfrac{1}{4z^2}$ if $z>1$

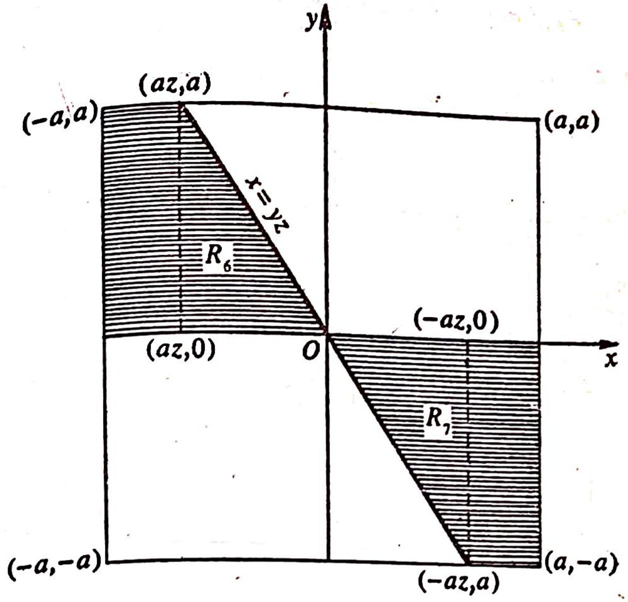


Fig. 6.10.24

Let $-1<z<0$.

In this case, $F_z(z)=\iint\limits_{R_6 \cup R_7} f(x,y)\,dx\,dy$

where R_6 and R_7 are the shaded regions shown in Fig. 6.10.24

$$=\int_{-a}^{az}\left\{\int_0^a f(x,y)\,dy\right\}dx+\int_{az}^{0}\left\{\int_{\frac{x}{z}}^{a} f(x,y)\,dy\right\}dx$$

$$+\int_0^{-az}\left\{\int_{\frac{x}{z}}^{0} f(x,y)\,dy\right\}dx+\int_{-az}^{a}\left\{\int_{-a}^{0} f(x,y)\,dy\right\}dx$$

$$=\frac{1}{4a^2}\left\{\int_{-a}^{az}a\,dx+\int_{az}^{0}\frac{x}{z}\,dx+\int_0^{-az}\left(-\frac{x}{z}\right)dx+\int_{-az}^{a}a\,dx\right\}$$

$$=\frac{1}{4a^2}\left\{a^2(z+1)-\frac{a^2z}{2}-\frac{a^2z}{2}+a^2(z+1)\right\}=\frac{1}{4}(z+2).$$

Hence we get

$$F_z'(z)=\tfrac{1}{4},\ -1<z<0$$

Let $z\leqslant -1$.

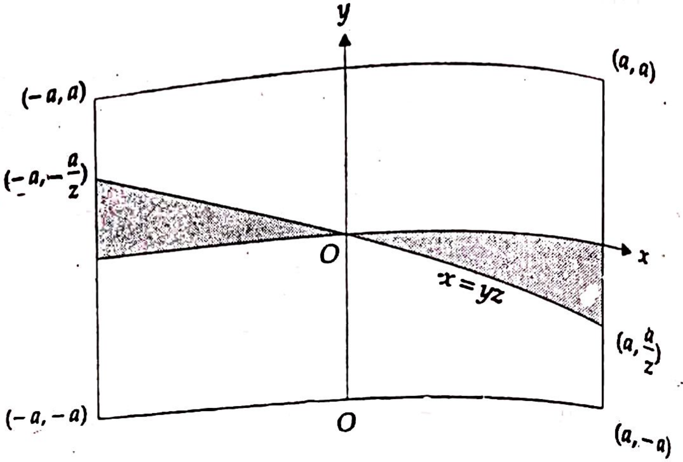


Fig. 6.10.25

Here we find that

$$F_z(z)=\frac{1}{4a^2}\left[-\frac{a^2}{2z}-\frac{a^2}{2z}\right]$$

or, $F_z(z)=-\dfrac{1}{4z}$ if $z\leqslant -1$

Hence, $F_z'(z)=\dfrac{1}{4z^2}$ if $z<-1$.

So the probability density function

$f_z(z)$ of $Z\left(=\dfrac{X}{Y}\right)$ is given by

$$f_z(z)=F_z'(z)=\frac{1}{4z^2}\ \text{if}\ z<-1\ \text{or}\ z>1$$

$$=\frac{1}{4}\ \text{if}\ -1<z<0\ \text{or}\ 0<z<1.$$

Ex. 26. *If X and Y are two independent random variables having density functions respectively as*

$$f_X(x)=e^{-x},\ x>0$$
$$=0\ ,\ \textit{elsewhere}$$

and $$f_Y(y)=e^{-y},\ y>0$$
$$=0\ ,\ \textit{elsewhere},$$

then find the distributions of (a) $X+Y$ *and* $\dfrac{X}{X+Y}$ (b) $X-Y$.

(a) We put $U=X+Y,\ V=\dfrac{X}{X+Y}$.

In terms of real variables, $u=x+y,\ v=\dfrac{x}{x+y}$.

$\therefore\ u=x+y$ and $x=uv$

that is, $x=uv,\ y=u(1-v)$.

$$\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix}=-u<0 \text{ for all } x>0,\ y>0.$$

X and Y being independent, the density function of (X, Y) is

$$f_{X,Y}(x,y)=e^{-(x+y)},\ x>0,\ y>0.$$

Now as the point (x, y) varies in the first quadrant of the xy plane, the point (u, v) varies in the region R of the first quadrant of the uv-plane shown in the shaded portion of Fig. 6.10.26.

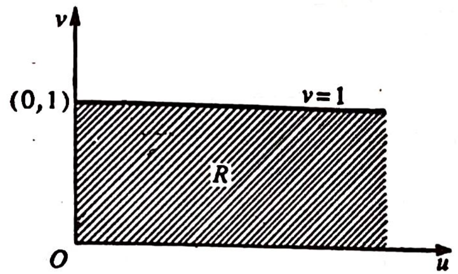


Fig. 6.10.26

Hence, the density function of the random variable (U, V) is given by

$$f_{U,V}(u,v)=e^{-(x+y)}(+u)=ue^{-u},\ 0<u<\infty,\ 0<v<1.$$

Hence, the density functions of $U=X+Y$ and $V=\dfrac{X}{X+Y}$ are respectively given by

$$f_U(u)=\int_0^1 ue^{-u}\,dv=ue^{-u},\ 0<u<\infty$$

and $$f_V(v)=\int_0^\infty ue^{-u}\,du=1,\quad 0<v<1.$$

(b) Let $U=X+Y, W=X-Y$.

In terms of real variables, $u=x+y,\ w=x-y$.

$$\therefore\ \frac{\partial(u,w)}{\partial(x,y)}=\begin{vmatrix}1 & 1\\ 1 & -1\end{vmatrix}=-2<0 \text{ for all } x, y.$$

The density function of the random variable (U, W) is given by,

$$f_{U,V}(u,w)=\tfrac{1}{2}e^{-u},\ 0<u<\infty,\ -u<w<u.$$

Hence, the density function of the random variable $W=X-Y$ is given by

$$f_W(w)=\int_w^\infty \tfrac{1}{2}e^{-u}\,du=\tfrac{1}{2}e^{-w},\text{ when } w>0$$

$$=\int_{-w}^\infty \tfrac{1}{2}e^{-u}\,du=\tfrac{1}{2}e^{w},\text{ when } w<0.$$

Ex. 27. *If X and Y are independent γ variates with parameters l and m respectively, then find the distributions of*

(a) $X+Y$ *and* $\dfrac{X}{X+Y}$, *(b)* $\dfrac{X}{Y}$.

(a) As X and Y are independent, the density function of (X, Y) is given by

$$f_{X,Y}(x,y)=\frac{e^{-(x+y)}x^{l-1}y^{m-1}}{\Gamma(l)\Gamma(m)},\ x>0,\ y>0.$$

We put $U=X+Y,\ V=\dfrac{X}{X+Y}$.

In terms of real variables, $u=x+y,\ v=\dfrac{x}{x+y}$

i.e. $x=uv,\ y=u(1-v)$.

As (x, y) varies in the first quadrant of the xy-plane, (u, v) varies in the region of the first quadrant of the uv-plane bounded by the lines $u=0,\ v=0,\ v=1$ (Fig. 6.10.26).

Also, $\dfrac{\partial(x,y)}{\partial(u,v)}=-u<0$ for all x, y.

$\therefore$ the density function of (U, V) is given by

$$f_{U,V}(u,v)=\frac{e^{-u}(uv)^{l-1}u^{m-1}(1-v)^{m-1}}{\Gamma(l)\Gamma(m)}u$$

$$=\frac{1}{\Gamma(l)\Gamma(m)}e^{-u}u^{l+m-1}v^{l-1}(1-v)^{m-1},$$

$$0<u<\infty,\ 0<v<1.$$

Hence, the marginal density function of $U=X+Y$ is

$$f_U(u)=\frac{1}{\Gamma(l)\Gamma(m)}e^{-u}u^{l+m-1}\int_0^1 v^{l-1}(1-v)^{m-1}\,dv$$

$$=\frac{B(l,m)}{\Gamma(l)\Gamma(m)}e^{-u}u^{l+m-1}$$

$$=\frac{e^{-u}u^{l+m-1}}{\Gamma(l+m)},\ 0<u<\infty,$$

and the marginal density function of $V=\dfrac{X}{X+Y}$ is given by

$$f_V(v)=\frac{v^{l-1}(1-v)^{m-1}}{\Gamma(l)\Gamma(m)}\int_0^\infty e^{-u}u^{l+m-1}\,du$$

$$=\frac{\Gamma(l+m)}{\Gamma(l)\Gamma(m)}v^{l-1}(1-v)^{m-1}$$

$$=\frac{v^{l-1}(1-v)^{m-1}}{B(l,m)},\ 0<v<1$$

which shows that U is a $\gamma(l+m)$ variate and V is a $\beta_1(l, m)$ variate.

(b) We put $U=X+Y,\ V=\frac{X}{Y}$.

In terms of real variables, $u=x+y,\ v=\frac{x}{y}$.

$$\therefore\ x=\frac{uv}{1+v},\ y=\frac{u}{1+v}.$$

$$\therefore\ \frac{\partial(u,v)}{\partial(x,y)}=-\frac{(1+v)^2}{u}<0 \text{ for all } x, y.$$

As x and y vary from 0 to ∞, u and v both vary from 0 to ∞ also.

$\therefore$ the density function of the random variable (U, V) is then given by

$$f_{U,V}(u,v)=\frac{e^{-u}\,u^{l+m-1}\,v^{l-1}}{\Gamma(l)\,\Gamma(m)(1+v)^{l+m}}.$$

If $f_V(v)$ be the marginal density function of V, it is then given by

$$f_V(v)=\frac{v^{l-1}}{\Gamma(l)\,\Gamma(m)(1+v)^{l+m}}\int_0^\infty e^{-u}\,u^{l+m-1}\,du$$

$$=\frac{\Gamma(l+m)}{\Gamma(l)\,\Gamma(m)}\ \frac{v^{l-1}}{(1+v)^{l+m}}$$

$$=\frac{v^{l-1}}{B(l,m)(1+v)^{l+m}},\ 0<v<\infty,$$

$[\because\ B(l,m)\,\Gamma(l+m)=\Gamma(l)\,\Gamma(m)]$

and this shows that $V=\frac{X}{Y}$ is a $\beta_2(l,m)$ variate.

Ex. 28. *If X and Y are two independent normal variates with parameters (m_x,σ_x) and (m_y,σ_y) respectively, then prove that $U=X+Y$ is a normal (m,σ) variate, where $m=m_x+m_y$, $\sigma^2=\sigma_x^2+\sigma_y^2$.*

The density function of (X, Y) is

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\,e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}\cdot e^{-\frac{(y-m_y)^2}{2\sigma_y^2}},\ -\infty<x,y<\infty,$$

We put $U=X+Y,\ V=X$.

In terms of real variables, $u=x+y,\ v=x$.

$$\therefore\ \frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix}1 & 1\\ 1 & 0\end{vmatrix}=-1<0 \text{ for all } x, y.$$

As x and y vary from $-\infty$ to ∞, u and v also vary from $-\infty$ to ∞,

$\therefore$ the density function of (U, V) is given by,

$$f_{U,V}(u,v)=\frac{1}{2\pi\sigma_x\sigma_y}\,e^{-\frac{(v-m_x)^2}{2\sigma_x^2}}\cdot e^{-\frac{(u-v-m_y)^2}{2\sigma_y^2}}$$

Now, $\frac{(v-m_x)^2}{2\sigma_x^2}+\frac{(u-v-m_y)^2}{2\sigma_y^2}$

$$=\frac{(v-m_x)^2}{2\sigma_x^2}+\frac{(u-v-m+m_x)^2}{2\sigma_y^2},\quad \because\ m=m_x+m_y$$

$$=\frac{(v-m_x)^2}{2\sigma_x^2}\left(1+\frac{\sigma_x^2}{\sigma_y^2}\right)+\frac{(u-m)^2}{2\sigma_y^2}-\frac{2(u-m)(v-m_x)}{2\sigma_y^2}$$

$$=\frac{\sigma^2}{\sigma_y^2}\ \frac{(v-m_x)^2}{2\sigma_x^2}-\frac{2(u-m)(v-m_x)}{2\sigma_y^2}+\frac{(u-m)^2}{2\sigma_y^2},$$

$$\because\ \sigma^2=\sigma_x^2+\sigma_y^2$$

$$=\frac{\sigma^2}{2\sigma_x^2\sigma_y^2}\left\{(v-m_x)^2-2(v-m_x)\frac{\sigma_x^2}{\sigma^2}(u-m)+\frac{\sigma_x^4}{\sigma^4}(u-m)^2\right\}$$

$$-\frac{\sigma_x^2}{2\sigma^2\sigma_y^2}(u-m)^2+\frac{(u-m)^2}{2\sigma_y^2}$$

$$=\frac{\sigma^2}{2\sigma_x^2\sigma_y^2}\left\{v-m_x-\frac{\sigma_x^2}{\sigma^2}(u-m)\right\}^2+\frac{(u-m)^2}{2\sigma_y^2\sigma^2}(\sigma^2-\sigma_x^2)$$

$$=\frac{\sigma^2}{2\sigma_x^2\sigma_y^2}\left\{v-m_x-\frac{\sigma_x^2}{\sigma^2}(u-m)\right\}^2+\frac{(u-m)^2}{2\sigma^2}$$

Hence, $f_{U,V}(u,v)$

$$=\frac{1}{2\pi\sigma_x\sigma_y}\ e^{-\frac{\sigma^2}{2\sigma_x^2\sigma_y^2}\left\{v-m_x-\frac{\sigma_x^2}{\sigma^2}(u-m)\right\}^2-\frac{(u-m)^2}{2\sigma^2}}.$$

Hence, the marginal density function of U is given by,

$$f_U(u)=\frac{e^{-\frac{(u-m)^2}{2\sigma^2}}}{2\pi\sigma_x\sigma_y}\int_{-\infty}^{\infty}e^{\frac{-\frac{\sigma^2}{2}\left\{v-m_x-\frac{\sigma_x^2}{\sigma^2}(u-m)\right\}^2}{\sigma_x^2\sigma_y^2}}\,dv$$

$$=\frac{e^{-\frac{(u-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}\cdot\frac{\sigma}{\sqrt{2\pi}\,\sigma_x\sigma_y}\int_{-\infty}^{\infty}e^{\frac{-\left\{v-m_x-\frac{\sigma_x^2}{\sigma^2}(u-m)\right\}^2}{2\frac{\sigma_x^2\sigma_y^2}{\sigma^2}}}\,dv$$

or, $$f_U(u) = \frac{e^{-\frac{(u-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma} \; \frac{1}{\sqrt{2\pi}\,\sigma'} \int_{-\infty}^{\infty} e^{-\frac{(v-m')^2}{2\sigma'^2}} \, dv$$

[where $\sigma' = \frac{\sigma_x \sigma_y}{\sigma}$, $m' = m_x + \frac{\sigma_x^2}{\sigma^2}(u-m)$.]

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(u-m)^2}{2\sigma^2}}, \; -\infty < u < \infty.$$

since for a normal (m', σ') distribution, the value of the last integral is 1. Thus,

$$f_U(u) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(u-m)^2}{2\sigma^2}}, \; -\infty < u < \infty,$$

which shows that U is normal (m, σ) variate.

Ex. 29. *If X_1, X_2 are independent random variables each having the density function $2xe^{-x^2}$ $(0 < x < \infty)$, then find the density function for the random variable $\sqrt{X_1^2 + X_2^2}$.*

X_1, X_2 being independent, the density function of (X_1, X_2) is

$$f_{X_1, X_2}(x_1, x_2) = 4x_1 x_2 e^{-(x_1^2 + x_2^2)}, \; 0 < x_1, x_2 < \infty.$$

We put $U = \sqrt{X_1^2 + X_2^2}$, $V = \frac{X_1}{X_2}$.

As x_1, x_2 vary from 0 to ∞, u and v also vary from 0 to ∞.

In terms of real variables, $u = \sqrt{x_1^2 + x_2^2}$, $v = \frac{x_1}{x_2}$.

$$\therefore \; \frac{\partial(u, v)}{\partial(x_1, x_2)} = \begin{vmatrix} \frac{x_1}{u} & \frac{x_2}{u} \\ \frac{1}{x_2} & -\frac{x_1}{x_2^2} \end{vmatrix} = -\frac{1}{u}\left(1 + \frac{x_1^2}{x_2^2}\right)$$

$$= -\frac{1+v^2}{u} < 0 \text{ for all } u, v.$$

$\therefore$ the density function of (U, V) is

$$f_{U,V}(u, v) = 4\frac{u^2 v}{1+v^2} e^{-u^2} \cdot \frac{u}{1+v^2}$$

$$\left[\because \; x_2 = \frac{u}{\sqrt{1+v^2}}, \; x_1 = \frac{uv}{\sqrt{1+v^2}}\right]$$

$$= \frac{4u^3 v}{(1+v^2)^2} e^{-u^2}, \; 0 < u, v < \infty.$$

Hence the marginal density function of U is

$$f_U(u) = u^3 e^{-u^2} \int_0^{\infty} \frac{4v \, dv}{(1+v^2)^2}$$

$$= 2u^3 e^{-u^2}, \; 0 < u < \infty,$$

Ex. 30. *The joint density function of the random variables X, Y is given by*

$$f_{X,Y}(x, y) = x + y, \; 0 < x < 1, \; 0 < y < 1$$
$$= 0, \text{ elsewhere.}$$

Find the distribution of (a) $X+Y$ and (b) XY.

(a) We put $U = X + Y$, $V = X$.

In terms of real variables, $u = x + y$, $v = x$

$$x = v, \; y = u - v.$$

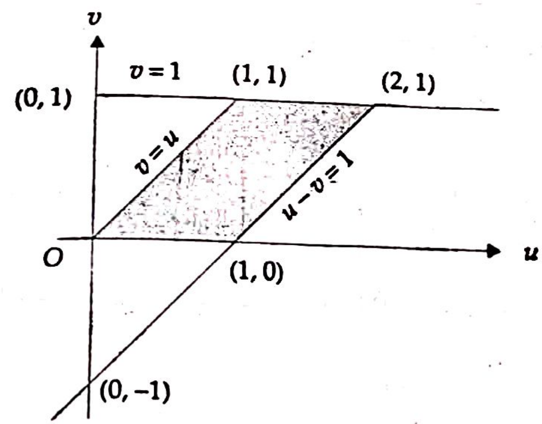


Fig. 6.10.27

As x, y vary from 0 to 1, u varies from 0 to 2 and v varies from 0 to 1.

$$\therefore \; \frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1 < 0 \text{ for all } x, y.$$

The joint density function of U and V is then

$$f_{U,V}(u, v) = (v + u - v) \, . \, 1$$
$$= u, \; 0 < v < 1, \; 0 < u - v < 1.$$

Then the marginal density function of U is given by

$$f_U(u)=\int_{-\infty}^{\infty} f_{U,V}(u,v)\,dv.$$

Let $0<u<1$. Then $f_U(u)=\int_0^u u\,dv=u^2$.

Let $1<u<2$. Then $f_U(u)=\int_{u-1}^{1} u\,dv=u(2-u)$.

Hence, the density function of $U=X+Y$ is given by

$$f_U(u)=\begin{cases} u^2 & 0<u<1 \\ u(2-u), & 1<u<2. \end{cases}$$

(b) We put $U=XY, V=X$.

In terms of real variables, $u=xy$, $v=x$.

$\therefore\ x=v,\ y=\dfrac{u}{v}$.

$$\therefore\ \frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix} 0 & 1 \\ \frac{1}{v} & -\frac{u}{v^2} \end{vmatrix}=-\frac{1}{v}<0 \text{ for all } u, v.$$

As x, y vary from 0 to 1, u and v also vary from 0 to 1.

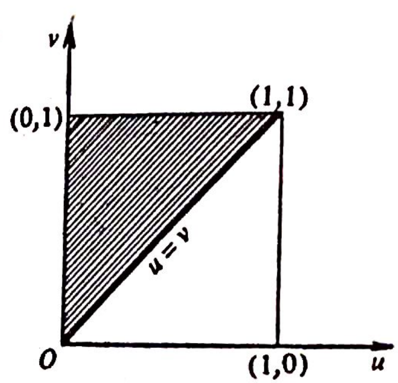


Fig. 6.10.28

The joint density function of U and V is

$$f_{U,V}(u,v)=\frac{1}{v}\left(v+\frac{u}{v}\right),\ 0<u<1,\ 0<\frac{u}{v}<1,$$

$$i.e.,\quad 0<u<v<1.$$

The marginal density function of U is then given by,

$$f_U(u)=\int_{-\infty}^{\infty}\left(1+\frac{u}{v^2}\right)dv=\int_u^1\left(1+\frac{u}{v^2}\right)dv$$

$$=2(1-u),\ 0<u<1.$$

Ex. 31. *Let (X, Y) have the general two-dimensional normal distribution, and we make a linear transformation*

$$U=(X-m_x)\cos\alpha+(Y-m_y)\sin\alpha$$
$$V=-(X-m_x)\sin\alpha+(Y-m_y)\cos\alpha,$$

where m_x, m_y, σ_x, σ_y, ρ have their usual meaning.

Show that U, V will be independent if

$$\tan 2\alpha=\frac{2\rho\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}.$$

We put $U=(X-m_x)\cos\alpha+(Y-m_y)\sin\alpha$

and $V=-(X-m_x)\sin\alpha+(Y-m_y)\cos\alpha$.

In terms of real variables,

$$u=(x-m_x)\cos\alpha+(y-m_y)\sin\alpha$$
$$v=-(x-m_x)\sin\alpha+(y-m_y)\cos\alpha.$$

Solving, $x=m_x+u\cos\alpha-v\sin\alpha$,

$$y=m_y+u\sin\alpha+v\cos\alpha.$$

$$\therefore\ \frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{vmatrix}=1>0 \text{ for all } x, y$$

Hence, the probability density function of (U, V) is given by

$$f_{U,V}(u,v)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(u\cos\alpha-v\sin\alpha)^2}{\sigma_x^2}+\frac{(u\sin\alpha+v\cos\alpha)^2}{\sigma_y^2}-2\rho\frac{(u\cos\alpha-v\sin\alpha)(u\sin\alpha+v\cos\alpha)}{\sigma_x\sigma_y}\right\}\right]$$

or,

$$f_{U,V}(u,v)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{u^2\left(\frac{\cos^2\alpha}{\sigma_x^2}+\frac{\sin^2\alpha}{\sigma_y^2}-\frac{\rho\sin 2\alpha}{\sigma_x\sigma_y}\right)\right.\right.$$

$$\left.\left.+v^2\left(\frac{\sin^2\alpha}{\sigma_x^2}+\frac{\cos^2\alpha}{\sigma_y^2}+\frac{\rho\sin 2\alpha}{\sigma_x\sigma_y}\right)+uv\left(-\frac{\sin 2\alpha}{\sigma_x^2}+\frac{\sin 2\alpha}{\sigma_y^2}-2\rho\frac{\cos 2\alpha}{\sigma_x\sigma_y}\right)\right\}\right]$$

$$=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{u^2}{2(1-\rho^2)}\left(\frac{\cos^2\alpha}{\sigma_x^2}+\frac{\sin^2\alpha}{\sigma_y^2}-\frac{\rho\sin 2\alpha}{\sigma_x\sigma_y}\right)\right\}$$

$$\times\exp\left\{-\frac{v^2}{2(1-\rho^2)}\left(\frac{\sin^2\alpha}{\sigma_x^2}+\frac{\cos^2\alpha}{\sigma_y^2}+\frac{\rho\sin 2\alpha}{\sigma_x\sigma_y}\right)\right\}$$

$$\times\exp\left\{-\frac{uv}{2(1-\rho^2)}\left(-\frac{\sin 2\alpha}{\sigma_x^2}+\frac{\sin 2\alpha}{\sigma_y^2}-2\rho\frac{\cos 2\alpha}{\sigma_x\sigma_y}\right)\right\}.$$

Now, if $\tan 2\alpha=\dfrac{2\rho\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}$

then $-\dfrac{\sin 2\alpha}{\sigma_x^2}+\dfrac{\sin 2\alpha}{\sigma_y^2}-2\rho\dfrac{\cos 2\alpha}{\sigma_x\sigma_y}=0$, so that the co-efficient of uv in the above expression for $f_{U,V}(u,v)$ vanishes.

Then $f_{U,V}(u,v)$ can be expressed as $f_{U,V}(u,v)=f_1(u)f_2(v)$, where $f_1(u)$ is a function of u only and $f_2(v)$ is a function of v only.

So the joint distribution function $F_{U,V}(u,v)$ of U and V is given by

$$F_{U,V}(u,v)=\int_{-\infty}^{u}\left\{\int_{-\infty}^{v}f_1(t)f_2(s)\,ds\right\}dt$$

$$=\int_{-\infty}^{u}\left[f_1(t)\int_{-\infty}^{v}f_2(s)\,ds\right]dt$$

$$=\left(\int_{-\infty}^{u}f_1(t)\,dt\right)\left(\int_{-\infty}^{v}f_2(s)\,ds\right)$$

So we get

$F_{U,V}(u,v)=F_1(u)\,F_2(v)$, for all u, v,

where $F_1(u)=\int_{-\infty}^{u}f_1(t)\,dt$ is a function of u only and $F_2(v)=\int_{-\infty}^{v}f_2(s)\,ds$ is a function of v only.

Hence, U, V are independent if $\tan 2\alpha=\dfrac{2\rho\sigma_x\sigma_y}{\sigma_x^2-\sigma_y^2}$.

Ex. 32. *Let X, Y be independent variates each having the density function ae^{-ax} $(0<x<\infty)$, where a is a positive constant. Find the density function of $\frac{X}{Y}$. Prove that the variate $\frac{Y}{X+Y}$ is uniformly distributed over $(0, 1)$.*

We put $U=\dfrac{X}{Y}$, $V=X$.

In terms of real variables, $u=\dfrac{x}{y}$, $v=x$; i.e., $x=v$, $y=\dfrac{v}{u}$.

The density function of (X, Y) is

$$f_{X,Y}(x,y)=a^2e^{-a(x+y)},\ 0<x, y<\infty.$$

Now $$\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}0 & 1\\ -\dfrac{v}{u^2} & \dfrac{1}{u}\end{vmatrix}=\frac{v}{u^2}>0 \text{ for all } x, y, >0.$$

As x, y vary from 0 to ∞, u, v both range from 0 to ∞.

The density function of (U, V) is then given by

$$f_{U,V}(u,v)=f_{X,Y}(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|$$

$$=a^2e^{-\frac{a(1+u)v}{u}}\cdot\frac{v}{u^2}$$

$$=\frac{a^2v}{u^2}e^{-\frac{av(1+u)}{u}},\ 0<u, v<\infty.$$

The marginal density function of U is then given by

$$f_U(u)=\int_0^\infty\frac{a^2v}{u^2}e^{-\frac{av(1+u)}{u}}\,dv=\frac{a^2}{u^2}\underset{B\to\infty}{Lt}\int_0^B ve^{-\frac{av(1+u)}{u}}\,dv$$

$$=\frac{a^2}{u^2}\underset{B\to\infty}{Lt}\left\{\left(-v\frac{u}{a(1+u)}e^{-\frac{av(1+u)}{u}}\right)_0^B+\int_0^B\frac{u}{a(1+u)}e^{-\frac{av(1+u)}{u}}\,dv\right\}$$

or, $f_U(u)=\frac{a^2}{u^2}\underset{B\to\infty}{Lt}\left\{-\frac{uB}{a(1+u)}e^{-\frac{aB(1+u)}{u}}-\frac{u^2}{a^2(1+u)^2}\left(e^{-\frac{aB(1+u)}{u}}-1\right)\right\}$

$$=\frac{a^2}{u^2}\frac{u^2}{a^2(1+u)^2}=(1+u)^{-2},\ 0<u<\infty.$$

To find the distribution of $\frac{Y}{X+Y}$, we put

$$Z=\frac{Y}{X+Y}=\frac{1}{1+\frac{X}{Y}}=\frac{1}{1+U}$$

In terms of real variables, $z=\frac{1}{1+u}$.

As u varies from 0 to ∞, z varies from 0 to 1

$$\therefore\ \frac{dz}{du}=-\frac{1}{(1+u)^2}<0 \text{ for all } u\in(0,\ \infty).$$

Hence, the density function of Z is given by

$$f_Z(z)=f_U(u)\left|\frac{du}{dz}\right|=(1+u)^{-2}(1+u)^2$$

$$=1,\ 0<z<1.$$

Hence, $Z=\frac{Y}{X+Y}$ is uniformly distributed in the interval (0, 1).

Ex. 33. *Let X and Y be two independent standard normal variates. Find the probability density function of $\sqrt{X^2+Y^2}$.*

Since X, Y are standard normal variates, $\frac{1}{2}X^2$, $\frac{1}{2}Y^2$ are $\gamma(\frac{1}{2})$ variates.

Also $\frac{1}{2}X^2$, $\frac{1}{2}Y^2$ are independent since X, Y are independent. Then by Ex. 27 (*a*) we find that $\frac{1}{2}(X^2+Y^2)$ is a $\gamma(1)$ variate.

Let $U=\sqrt{X^2+Y^2}$.

Then $U=\sqrt{2V}$ where $V=\frac{1}{2}(X^2+Y^2)$ is a $\gamma(1)$ variate.

In real variables we have

$$u=\sqrt{2v}.$$

$$\therefore\ \frac{du}{dv}=\frac{1}{\sqrt{2v}}>0 \text{ for } v>0.$$

So the p.d.f $f_U(u)$ of U is given by

$$f_U(u)=\sqrt{2v}\,f_V(v)$$

where the p.d.f $f_V(v)$ of V is given by

$$f_V(v)=e^{-v},\text{ for } 0<v<\infty,$$

Hence $f_U(u)=\sqrt{2v}\cdot e^{-v}$ if $v>0$

or, $f_U(u)=ue^{-\frac{u^2}{2}}$ if $u>0$.

So the probability density function of $U=\sqrt{X^2+Y^2}$ is given by

$$f_U(u)=ue^{-\frac{u^2}{2}},\ u>0.$$

Ex. 34. *The random variables X and Y are independent and their probability density functions are respectively given by*

$$f(x)=\frac{1}{\pi}\frac{1}{\sqrt{1-x^2}},\ |x|<1 \text{ and } g(y)=ye^{-\frac{y^2}{2}},\ y>0.$$

Find the joint density functions of Z and W, where $Z=XY$ and $W=X$. Deduce the probability density function of Z.

We put $Z=XY$ and $W=X$.

In terms of real variables, $z=xy$, $w=x$;

that is, $x=w$, $y=\frac{z}{w}$.

Therefore $\frac{\partial(z,w)}{\partial(x,y)}=\begin{vmatrix}\frac{\partial z}{\partial x} & \frac{\partial z}{\partial y}\\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y}\end{vmatrix}=\begin{vmatrix} y & x\\ 1 & 0\end{vmatrix}=-x=-w$

which changes sign

So we proceed as follows :

Let $F_{Z,W}(z,w)$ be the joint distribution function of Z and W.

Then $F_{Z,W}(z,w)=P(XY\leqslant z,\ W\leqslant w)$

$$=P(XY\leqslant z,\ X\leqslant w).$$

Case I. Let $z > 0$.

The joint p.d.f. of X and Y is given by

$$f_{X,Y}(x, y)=\frac{1}{\pi\sqrt{1-x^2}}\cdot ye^{-\frac{y^2}{2}},$$

for $-1 < x < 1, 0 < y < \infty$.

[$\because$ X, Y are independent.]

In this case, if $w < 0$ then the event $(XY \leqslant z, X \leqslant w)$ happens if and only if $(X \leqslant w)$ happens and so $P(XY \leqslant z, X \leqslant w)$

$$=P(X \leqslant w) \text{ if } w \leqslant 0$$

$$=\int_{-1}^{w}\frac{1}{\pi\sqrt{1-x^2}}\,dx \quad \text{if } -1 < w < 0$$

$$=\frac{1}{\pi}\left[\sin^{-1} w+\frac{\pi}{2}\right] \quad \text{if } -1 < w < 0.$$

Also we see that $P(X \leqslant w)=0$ if $w \leqslant -1$.

So $F_{Z,W}(z, w)=0$ if $w \leqslant -1$

$$=\frac{1}{2}+\frac{1}{\pi}\sin^{-1} w \quad \text{if } -1 < w < 0.$$

Now, if $0 \leqslant w < 1$ $(Z > 0)$ we see that $(XY \leqslant z, X \leqslant w)$ happens if and only if the random point (X, Y) lies in the shaded region of Fig. 6.10.29.

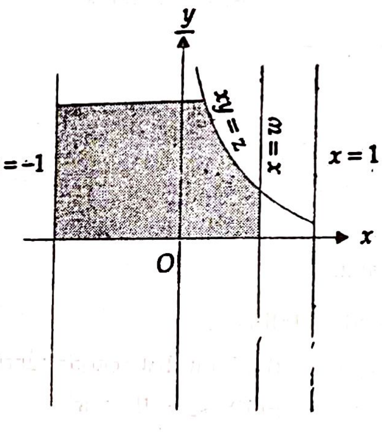


Fig. 6.10.29

Hence, if $z > 0$ and $0 \leqslant w < 1$

then $$F_{Z,W}(z, w)=\int_{-1}^{0}\left\{\frac{1}{\pi\sqrt{1-x^2}}\int_0^\infty ye^{-\frac{y^2}{2}}\,dy\right\}dx$$

$$+\int_0^w\left\{\int_0^{\frac{z}{x}}\frac{1}{\pi\sqrt{1-x^2}}ye^{-\frac{y^2}{2}}\,dy\right\}dx,$$

So $$F_{Z,W}(z, w)=\frac{1}{2}+\int_0^w\frac{1}{\pi\sqrt{1-x^2}}\left[1-e^{-\frac{z^2}{2x^2}}\right]dx$$

if $z > 0, 0 \leqslant w < 1$.

Further we see that if $z > 0, w \geqslant 1$,

then $$F_{Z,W}(z, w)=\frac{1}{2}+\int_0^1\frac{1}{\pi\sqrt{1-x^2}}\left[1-e^{-\frac{z^2}{2x^2}}\right]dx.$$

Thus in case I $(z > 0)$ it is proved that

$$F_{Z,W}(z, w)=0 \quad \text{if } w \leqslant -1$$

$$=\frac{1}{2}+\frac{1}{\pi}\sin^{-1}w \quad \text{if } -1 < w < 0$$

$$=\frac{1}{2}+\int_0^w\frac{1}{\pi\sqrt{1-x^2}}\left[1-e^{-\frac{z^2}{2x^2}}\right]dx, \text{ if } 0 \leqslant w < 1,$$

$$=\frac{1}{2}+\int_0^1\frac{1}{\pi\sqrt{1-x^2}}\left[1-e^{-\frac{z^2}{2x^2}}\right]dx, \text{ if } w \geqslant 1,$$

If $f_{Z,W}(z, w)$ be the joint p.d.f. of Z and W then

$f_{Z,W}(z, w)=\dfrac{\partial^2 F_{Z,W}(z, w)}{\partial z\, \partial w}$ is given in case I $(z > 0)$ by

$$f_{Z,W}(z, w)=0 \text{ if } w <-1 \text{ or } -1 < w < 0$$

$$=\frac{1}{\pi\sqrt{1-w^2}}\cdot\frac{z}{w^2}e^{-\frac{z^2}{2w^2}}, \text{ if } 0 < w < 1.$$

$=0$ if $w < 1$.

So if $z > 0$, the p.d.f. $f_z(z)$ of Z is given by

$$f_z(z)=\frac{z}{\pi}\int_0^1 \frac{1}{w^2\sqrt{1-w^2}}\, e^{-\frac{z^2}{2w^2}}\, dw.$$

Now using the substitution $\frac{w}{\sqrt{1-w^2}}=u$ we find that

$$f_z(z)=\frac{z}{\pi} e^{-\frac{z^2}{2}}\int_0^\infty \frac{1}{u^2}\, e^{-\frac{z^2}{2u^2}}\, du$$

$$=\frac{z}{\pi} e^{-\frac{z^2}{2}}\int_0^\infty \frac{\sqrt{2}}{z}\, e^{-v^2}\, dv$$

$$\left(v=\frac{z}{u\sqrt{2}},\ z>0\right)$$

$$=\left(\frac{\sqrt{2}}{\pi}\int_0^\infty e^{-v^2}\, dv\right)e^{-\frac{z^2}{2}}$$

$$=\frac{\sqrt{2}}{\pi}\cdot\frac{1}{2}\sqrt{\pi}\, e^{-\frac{z^2}{2}}$$

$$=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}},\ \text{if } z>0.$$

It can be shown similarly that

$$f_z(z)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}},\ \text{if } z<0.$$

Since $f_z(z)$ can be defined arbitrarily at $z=0$, we take

$$f_z(0)=\frac{1}{\sqrt{2\pi}}.$$

Then $f_z(z)=\frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}},\ -\infty < z < \infty$ and this shows that $Z=XY$ has normal distribution.

Ex. 35. *If X and Y are independent normal (0, 1) random variables and $(R, \odot)$ is the representation in polar co-ordinates of the point (X, Y) in the cartesian plane, then find the distribution of R^2 and $\odot$.*

Here $X=R\cos\odot$ and $Y=R\sin\odot$.

In real variables, $x=r\cos\theta,\ y=r\sin\theta$.

Here $-\infty<x<\infty,\ -\infty<y<\infty$ and $0\leq r<\infty,\ 0\leq\theta<2\pi$.

Also, $\frac{\partial(x, y)}{\partial(r, \theta)}=r>0$, when $0<r<\infty$.

Since X and Y are independent standard normal variates the joint density function of R and $\odot$ is

$$f_{R,\odot}(r, \theta)=\frac{r}{2\pi}e^{-\frac{r^2}{2}}.$$

Hence, the marginal density function of R is given by

$$f_R(r)=\int_0^{2\pi} f_{R,\odot}(r, \theta)\, d\theta$$

$$=\frac{r}{2\pi}e^{-\frac{r^2}{2}}\int_0^{2\pi} d\theta=re^{-\frac{r^2}{2}},\ 0<r<\infty.$$

Again, the marginal density function of $\odot$ is given by

$$f_\odot(\theta)=\frac{1}{2\pi}\int_0^\infty re^{-\frac{r^2}{2}}\, dr=\frac{1}{2\pi},\ 0<\theta<2\pi.$$

Let $U=R^2$. In real variables $u=r^2$.

When $0<r<\infty,\ 0<u<\infty$.

$\therefore\ \frac{du}{dr}=2r>0$ when $0<r<\infty$.

$\therefore$ the density function of $U=R^2$ is given by

$$f_U(u)=f_R(r)\left|\frac{dr}{du}\right|=\frac{1}{2}e^{-\frac{u}{2}},\ 0<u<\infty.$$

Ex. 36. *X, Y are independent random variables having density functions given by*

$$f_X(x)=\frac{1}{x^2}, \text{ if } x \geqslant 1$$
$$=0, \text{ elsewhere}$$
$$\text{and } f_Y(y)=1, \text{ if } 0<y<1$$
$$=0, \text{ elsewhere},$$

respectively. Find the p.d.f. of $X+Y$.

Let $U=X+Y$, $V=X$.

In real variables, $u=x+y$, $v=x$.

Now, $\frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix}=-1$, which does not change sign. The probability density function of $(U\ V)$ is then given by

$$f_{U,V}(u,v)=\frac{1}{x^2}\cdot 1, \text{ if } x \geqslant 1, 0<y<1$$

that is, $f_{U,V}(u,v)=\frac{1}{v^2}$, if $v \geqslant 1, 0<u-v<1$

$=0$, elsewhere.

Then the probability density function of $U=X+Y$ is given by

$$f_U(u)=\int_1^\infty f_{U,V}(u,v)\,dv.$$

Now the region where $f_{U,V}(u,v)\neq 0$ is the region $R=\{(u,v):\ 0<u-v<1, v\geqslant 1\}$ in the $u-v$ plane, shaded in Fig. 6.10.30.

Let $1<u<2$. Then $f_U(u)=\int_1^u \frac{1}{v^2}\,dv=1-\frac{1}{u}$.

Let $2 \leqslant u<\infty$. Then $f_U(u)=\int_{u-1}^u \frac{1}{v^2}\,dv=-\frac{1}{u}+\frac{1}{u-1}=\frac{1}{u(u-1)}$.

Then the required density function of U is given by

$$f_U(u)=\begin{cases} 1-\frac{1}{u}, & \text{if } 1<u<2 \\ \frac{1}{u(u-1)}, & \text{if } u \geqslant 2. \end{cases}$$

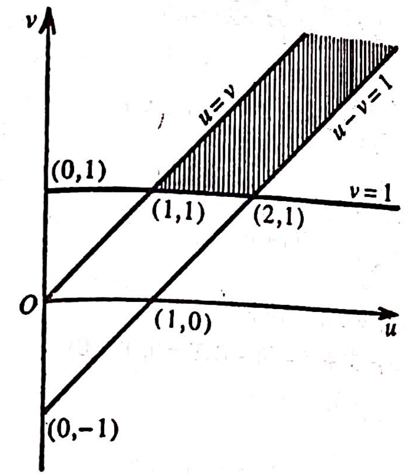


Fig. 6.10.30

Ex. 37. *If X, Y are independent Poisson variates with parameters μ_1, μ_2 respectively, then show that $X+Y$ is a Poisson variate with parameter $\mu_1+\mu_2$.*

Let $U=X+Y$.

The spectrum of U is the enumerable set $\{0, 1, 2, \ldots\}$. Here we have

$$P(X=i)=\frac{e^{-\mu_1}\mu_1^i}{i!}, \text{ for } i=0,1,2,\ldots\ldots$$

and $P(X=j)=\frac{e^{-\mu_2}\mu_2^j}{j!}$, for $j=0,1,2,\ldots\ldots$

Then for any given positive integer k,

$$P(U=k)=\sum_{i+j=k} P(X=i, Y=j)$$

$$=\sum_{i+j=k} P(X=i)\,P(Y=j), \text{ since } X \text{ and } Y \text{ are independent.}$$

or, $P(U=k)=\sum_{i=0}^{k} e^{-(\mu_1+\mu_2)} \dfrac{\mu_1^{\,i}\,\mu_2^{\,k-i}}{i!\,(k-i)!}$

$$=\frac{e^{-(\mu_1+\mu_2)}}{k!}\sum_{i=0}^{k}\frac{\mu_1^{\,i}\,\mu_2^{\,k-i}}{i!\,(k-i)!}k!$$

$$=\frac{e^{-(\mu_1+\mu_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\mu_1^{\,i}\mu_2^{\,k-i}$$

$$=\frac{e^{-(\mu_1+\mu_2)}}{k!}(\mu_2+\mu_1)^k.$$

Also for $k=0$, we find that

$$P(U=0)=P(X+Y=0)=P(X=0,\ Y=0)$$

$$=e^{-\mu_1}\cdot e^{-\mu_2}$$

$$=e^{-(\mu_1+\mu_2)}\cdot\frac{(\mu_1+\mu_2)^0}{0!}.$$

Thus it is proved that for any non-negative integer k,

$$P(U=k)=\frac{e^{-(\mu_1+\mu_2)}\cdot(\mu_1+\mu_2)^k}{k!}$$

and this shows that $U=X+Y$ is a Poisson variate with parameter $\mu_1+\mu_2$.

Ex 38. *If $X_1, X_2, \ldots, X_n$ are mutually independent and each X_i has uniform distribution over the interval (a, b), then find the density function of the random variable U, given by*

$$U=\min.\{X_1, X_2, \ldots, X_n\}.$$

Let $a<u<b$.

We observe that the event $(U>u)$ happens if and only if

$$(X_1>u,\ X_2>u, \ldots, X_n>u) \text{ happens.}$$

Then $P(U>u)=P(X_1>u,\ X_2>u,\ \ldots,\ X_n>u)$

$$=P(X_1>u)\,P(X_2>u)\ldots\ldots P(X_n>u),$$

since $X_1, X_2, \ldots X_n$ are independent

$$\left(\int_u^b \frac{dx_1}{b-a}\right)\left(\int_u^b \frac{dx_2}{b-a}\right)\cdots\left(\int_u^b \frac{dx_n}{b-a}\right)=\frac{(b-u)^n}{(b-a)^n}.$$

Let $F_U(u)$ be the distribution function of U. Then

$$F_U(u)=P(U\leq u)=1-P(U>u)$$

$$=1-\frac{(b-u)^n}{(b-a)^n},\ \text{if } a<u<b.$$

If $u>b$, $F_U(u)=P(U\leq u)=1$.

If $u\leq a$, $F_U(u)=P(U\leq u)=0$.

So, if $f_U(u)$ be the density function of U, then

$$f_U(u)=F'_U(u)=\begin{cases}\dfrac{n(b-u)^{n-1}}{(b-a)^n}, & \text{if } a<u<b\\ 0, & \text{if } u<a \text{ or } u>b\end{cases}$$

and $f_U(a)$ is undefined since $LF_U'(a)\neq RF_U'(a)$.

Ex 39. *Let $X_1, X_2, \ldots, X_n$ be mutually independent discrete random variables each having p.m.f. given by $P(X_i=k)=\dfrac{1}{N}$ for $k=1, 2, \ldots, N$, where $i\in\{1, 2, \ldots, n\}$. Find the probability mass function of U_n, where $U_n=\min\{X_1, X_2, \ldots, X_n\}$.* [C. H. (Math.) '91]

Here the spectrum of U_n is the set $\{1, 2, \ldots, N\}$.

We see that the event $(U_n\geq k)$ happens if and only if the event

$$(X_1\geq k,\ X_2\geq k, \ldots, X_n\geq k) \text{ happens.}$$

So $P(U_n\geq k)=P(X_1\geq k)P(X_2\geq k)\ldots P(X_n\geq k)$

$$=\frac{N-(k-1)}{N}\cdot\frac{N-(k-1)}{N}\cdots\frac{N-(k-1)}{N}$$

$$=\left(\frac{N-k+1}{N}\right)^n.$$

Similarly, we have $P(U_n\geq k+1)=\left(\dfrac{N-k}{N}\right)^n.$

Now we observe that the event $(U_n > k)$ can be expressed as

$$(U_n > k) = (U_n > k+1) + (U_n = k),$$

where $(U_n > k+1)$ and $(U_n = k)$ are two mutually exclusive events.

Therefore, $P(U_n > k) = P(U_n > k+1) + P(U_n = k)$

or, $\left(\dfrac{N-k+1}{N}\right)^n = \left(\dfrac{N-k}{N}\right)^n + P(U_n = k)$

or, $P(U_n = k) = \left(\dfrac{N-k+1}{N}\right)^n - \left(\dfrac{N-k}{N}\right)^n.$

Hence, the required p.m.f. is given by

$$P(U_n = k) = \left(\frac{N-k+1}{N}\right)^n - \left(\frac{N-k}{N}\right)^n, \text{ for } k = 1, 2, \ldots, N.$$

Ex. 40. *A rectangular bridge ABCD of width $AB = 20$ m. $(-10 \leqslant x \leqslant 10)$ and length $BC = 200$ m. $(-100 \leqslant y \leqslant 100)$ spans a river. In the artillery shelling of the bridge the hitting point (x, y) on the bridge is a pair of independent normal variates with standard deviations $\sigma_x = 10$ m.; $\sigma_y = 40$ m., the co-ordinates of the aiming point of the shelling being the expectation values (m_x, m_y) of the two random variables. Find the probability of hitting the bridge in a single shot when* [C. H. (Math.) '94]

(i) $(m_x, m_y) = (0, 0)$, (ii) $(m_x, m_y) = (5, 20)$.

Given $\phi(u) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^u exp\left(-\frac{t^2}{2}\right) dt$

u	0·5	1·0	1·5	2·0	2·5	3·0
$\phi(u)$	·1915	·3413	·4332	·4772	·4038	·4987.

Let $U = \dfrac{X - m_x}{\sigma_x}$, $V = \dfrac{Y - m_y}{\sigma_y}$.

Then U, V are independent standard normal variates. It is given that $\sigma_x = 10$, $\sigma_y = 40$.

Case (I) $m_x = 0$, $m_y = 0$.

Here $U = \dfrac{X}{10}$, $V = \dfrac{Y}{40}$.

The probability of hitting the bridge in a single shot is

$$P(-10 < X < 10, -100 < Y < 100)$$
$$P(-10 < X < 10)\, P(-100 < Y < 100)$$

since X and Y are independent

$$= P(-1 < U < 1)\, P(-2{\cdot}5 < V < 2{\cdot}5)$$
$$= 4P(0 < U < 1)\, P(0 < V < 2{\cdot}5)$$
$$= 4 \frac{1}{\sqrt{2\pi}} \int_0^1 \exp\left(-\frac{t^2}{2}\right) dt \cdot \frac{1}{\sqrt{2\pi}} \int_0^{2\cdot5} \exp\left(-\frac{t^2}{2}\right) dt$$
$$= 4\,\phi(1)\,\phi(2{\cdot}5) = 4 \times {\cdot}3413 \times {\cdot}4938 = {\cdot}67.$$

Case (II) $m_x = 5$, $m_y = 20$.

Here $U = \dfrac{X-5}{10}$, $V = \dfrac{Y-20}{40}$.

The required probability is

$$P(-10 < X < 10)\, P(-100 < Y < 100)$$
$$= P(-1{\cdot}5 < U < {\cdot}5)\, P(-3 < V < 2)$$
$$= \left[\frac{1}{\sqrt{2\pi}} \int_{-1\cdot5}^{0} \exp\left(-\frac{t^2}{2}\right) dt + \frac{1}{\sqrt{2\pi}} \int_0^{\cdot5} \exp\left(-\frac{t^2}{2}\right) dt\right]$$
$$\times \left[\frac{1}{\sqrt{2\pi}} \int_{-3}^{0} \exp\left(-\frac{t^2}{2}\right) dt + \frac{1}{\sqrt{2\pi}} \int_0^{2} \exp\left(-\frac{t^2}{2}\right) dt\right]$$
$$= \{\phi(1{\cdot}5) + \phi({\cdot}5)\}\{\phi(3) + \phi(2)\}$$
$$= ({\cdot}4332 + {\cdot}1915)({\cdot}4987 + {\cdot}4772)$$
$$= {\cdot}6247 \times {\cdot}9759 = {\cdot}61.$$

Ex. 41. *In the quadratic equation $x^2 + 2ax + b = 0$, a and b independently are equally to take any value in the interval $(-1, 1)$. Find the probability that the roots are real.* [C. H. (Math.) '85

Let A and B be the random variables corresponding to the real coefficients a and b of the given equation. From the given condition, A and B are independent and both uniformly distributed

in the interval $(-1, 1)$. Then the two-dimensional random variable (A, B) is uniformly distributed in the square

$$R=\{(x, y) : -1 < x < 1,\ -1 < y < 1\},$$

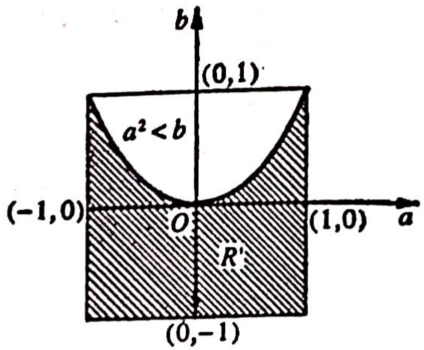


Fig. 6.10.31

its area is also denoted by R. Now the given equation will have real roots if and only if $4A^2-4B \geqslant 0$, *i.e.*, $A^2 \geqslant B$

or, if and only if (A, B) lies in the shaded region R' shown in Fig. 6.10.31, its area being also denoted by R'.

Now $R=4$, $R'=2+2\int_0^1 b\,da=2+2\int_0^1 a^2 da=\frac{8}{3}$.

Hence, by (6.7.3), the required probability $=\frac{R'}{R}=\frac{2}{3}$.

Ex. 42. *Two people agree to meet at a definite place between 12 and 1 O'clock with the understanding that each will wait 20 minutes for the other. What is the probability that they will meet?*

Let X and Y be the random variables corresponding to the time of arrival measured in hours of the two people from the instant 12 O'clock. X and Y are independent random variables both uniformly distributed in $(0, 1)$. Then the two-dimensional random variable (X, Y) is uniformly distributed over the square region $R=\{(x, y) : 0 < x < 1, 0 < y < 1\}$. Now the meeting will take place if $|X-Y| < \frac{1}{3}$, that is, if $X \geqslant Y$ and $X-Y \leqslant \frac{1}{3}$, or, $X < Y$ and $Y-X \leqslant \frac{1}{3}$.

In other words the event that 'the meeting will take place' will take place if and only if (X, Y) lies in the shaded region R' (Fig. 6.10.32) where

$$R'=\{(x, y) : x > y \text{ and } x-y < \tfrac{1}{3} \text{ or } x < y \text{ and } y-x < \tfrac{1}{3}\}.$$

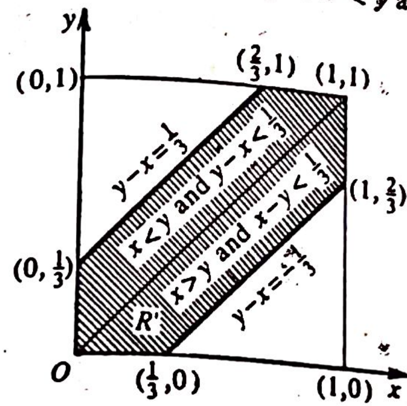


Fig. 6.10.32

We denote the corresponding areas by R and R'.

Now $R=1$ and $R'=1-2(\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{2}{3})=\frac{5}{9}$.

$\therefore$ required probability $=\frac{R'}{R}=\frac{5}{9}$.

Ex. 43. *Buffon's Needle Problem: A vertical board is ruled with horizontal parallel lines at constant distance b apart. A needle of length a (< b) is thrown at random on the board. Find the probability that it will intersect one of the lines.*

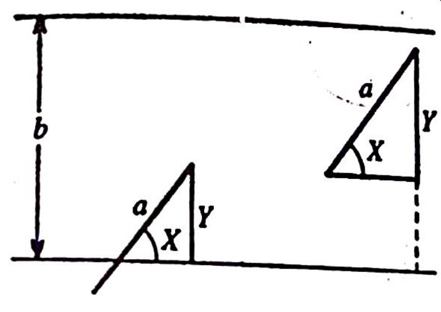


Fig. 6.10.33

We denote the inclination of the needle with horizontal by the random variable X and the distance of the upper end of the

needle from the nearest ruling below it by Y. Clearly then X is uniformly distributed in the interval $(0, \pi)$ and Y is uniformly distributed in the interval $(0, b)$. Now X and Y are independent random variables and distribution of the two-dimensional random variable (X, Y) is uniform in the rectangular region R given by

$$R = \{(x, y) : 0 < x < \pi, 0 < y < b\}.$$

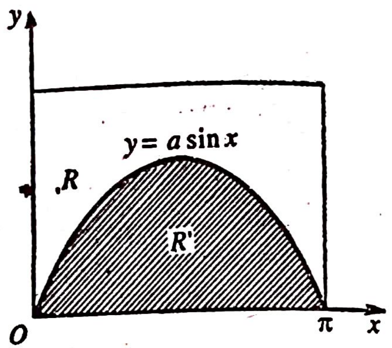


Fig. 6.10.34

Now the needle will intersect one of the horizontal lines if X and Y satisfy the inequality $0 \leqslant Y \leqslant a \sin X$, in other words, if (X, Y) lies in the region R' given by

$$R' = \{(x, y) : 0 \leqslant y \leqslant a \sin x\}.$$

Denoting the corresponding areas by R and R', required probability

$$= \frac{R'}{R} = \frac{\int_0^{\pi} a \sin x \, dx}{\pi b} = \frac{2a}{\pi b}.$$

Ex. 44. *Two numbers are independently chosen at random between 0 and 1. Show that the probability that their product is less than a constant k $(0 < k < 1)$ is $k(1 - \log k)$.*

Let the random variables X and Y denote the two numbers chosen. Then, as before, (X, Y) is uniformly distributed in the square region R given by

$$R = \{(x, y) : 0 < x < 1, 0 < y < 1\}.$$

Now the product of the two numbers will be less than k, if $XY < k$. Thus the given event will happen if and only if (X, Y) lies in the region

$$R' = \{(x, y) : xy < k, 0 < x < 1, 0 < y < 1\},$$

as shown in the shaded region of Fig. 6.10.35.

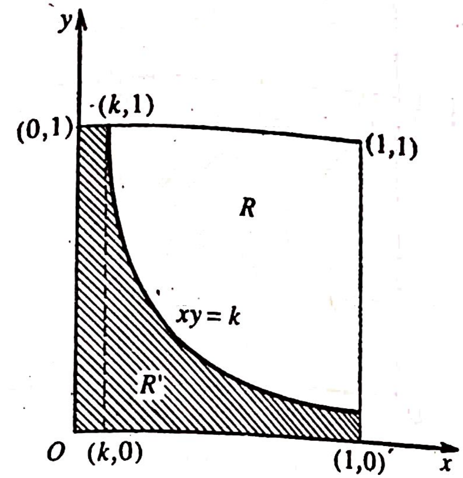


Fig. 6.10.35

We denote the corresponding areas by R and R' also.

Hence, the required probability

$$= \frac{R'}{R} = \frac{k \times 1 + \int_k^1 \frac{k}{x} dx}{1} = k(1 - \log k).$$

Ex. 45. *What is the probability that the sum of two numbers, chosen randomly from the interval $(0, 1)$, is greater than 1, while the sum of their squares is less than 1?*

Let X and Y be the random variables denoting the two numbers, chosen at random from the interval $(0, 1)$. Then, as before, the random variable (X, Y) is uniform in the square region R, given by

$$R = \{(x, y) : 0 < x < 1, 0 < y < 1\}.$$

Now the given event can be represented by joint inequalities '$X+Y>1$ and $X^2+Y^2<1$' or that (X, Y) lies in the region R' given by

$$R'=\{(x, y): x+y>1 \text{ and } x^2+y^2<1\},$$

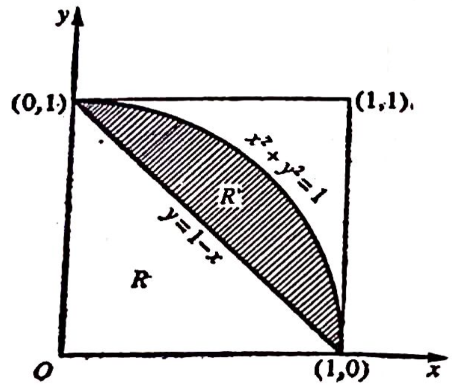


Fig. 6.10.36

which is shown in the shaded region of Fig. 6.10.36. As before, we denote the corresponding areas by R and R'. Hence the required probability

$$=\frac{R'}{R}=\frac{\frac{\pi}{4}-\frac{1}{2}}{1}=\frac{\pi}{4}-\frac{1}{2}.$$

Ex. 46. *A straight line AB is divided by a point C into two parts AC and CB whose lengths are a and b respectively $(a>b)$. If two points P and Q are independently chosen at random on AC and CB respectively, then find the probability that AP, PQ and QB can form a triangle.*

Let the random variables X, Y denote the distances AP and AQ respectively of the two points P and Q chosen at random from the segments AC and CB of lengths a and b $(a>b)$ respectively. Then as before, X and Y being independent, the random variable

(X, Y) is uniformly distributed in the rectangular region R, given by

$$R=\{(x, y): 0<x<a,\ a<y<a+b\}.$$

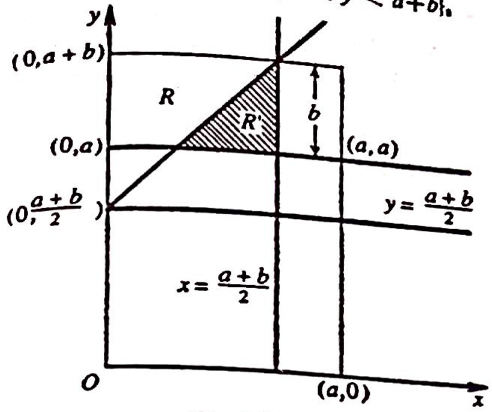


Fig. 6.10.37

Now $AP=X$, $PQ=Y-X$ and $QB=a+b-Y$.

Now AP, PQ and QB will form a triangle if

(i) $X+Y-X>a+b-Y$, i.e., $Y>\frac{a+b}{2}$, which is true as $Y>a$ and $a>\frac{a+b}{2}$ $(\because\ a>b)$,

(ii) $X+a+b-Y>Y-X$, i.e, $Y<\frac{a+b}{2}+X$

and (iii) $Y-X+a+b-Y>X$, i.e., $X<\frac{a+b}{2}$.

Now the above three inequalities hold, if the random variable (X, Y) lies in the region R' lying within R and bounded by the lines

$$y=a,\quad y=x+\frac{a+b}{2} \text{ and } x=\frac{a+b}{2},$$

as shown by the shaded portion in Fig. 6.10.37.

Hence the required probability

$$=\frac{R'}{R}=\frac{\frac{1}{2}b^2}{ab}=\frac{b}{2a}.$$

Ex. 47. *Two points P, Q are independently chosen at random on a circle and A is a fixed point also on the circle. Find the probability that the points A, P, Q will lie on the same semi-circle.*

Let the random variables X and Y denote the angles made by OP and OQ with OA respectively, O being the centre of the circle. Then as before, the random variable (X, Y) is uniformly distributed in the square region R, given by

$$R = \{(x, y) : 0 < x < 2\pi,\ 0 < y < 2\pi\}.$$

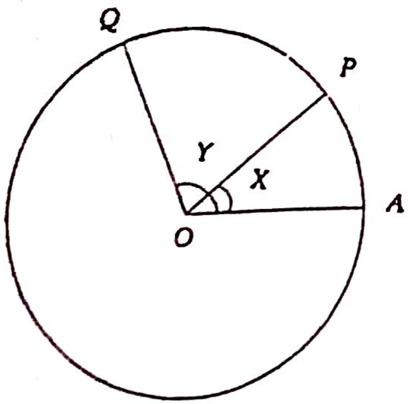


Fig. 6.10.38

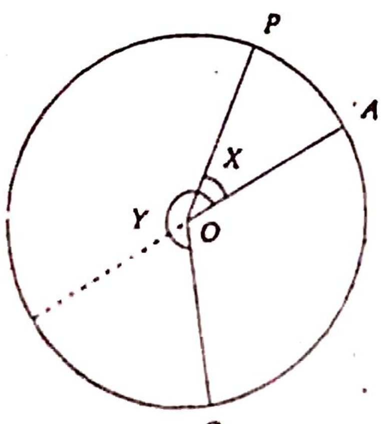


Fig. 6.10.39

Let $X < Y$. In this case, the three points will lie on the same semi-circle, if the following inequalities hold :

$$Y < \pi \quad \text{or} \quad Y > X + \pi.$$

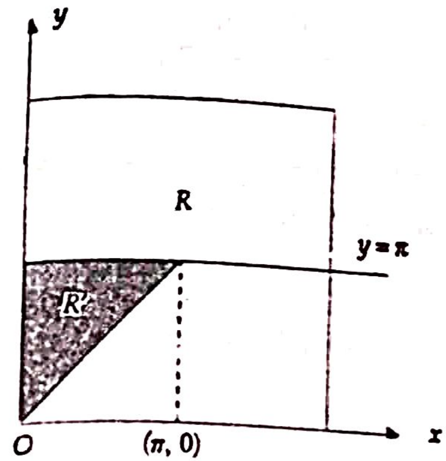


Fig. 6.10.40

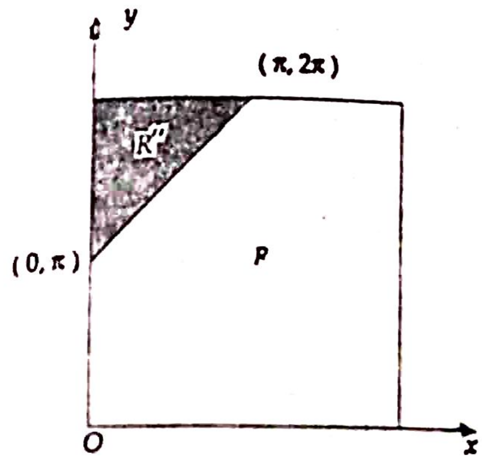


Fig. 6.10.41

In other words, the given event means that the random variable (X, Y) lies in the triangular region R' bounded by $x=0$, $y=\pi$, $x=y$ (Fig. 6.10.40) or lies in the triangular region R'' bounded by the lines $x=0$, $y=2\pi$, $y=x+\pi$ (Fig. 6.10.41).

The area of each of R' and R'' is $\frac{\pi^2}{2}$. Now considering the case $X > Y$ and from symmetry we find that the required probability

$$=\frac{2\left(\frac{\pi^2}{2}+\frac{\pi^2}{2}\right)}{4\pi^2}=\frac{1}{2}.$$

Ex. 48. *Prove that (with usual notations),*

$$F_X(x)+F_Y(y)-1 \leqslant F_{X,Y}(x, y) \leqslant \sqrt{F_X(x)\,F_Y(y)}, \text{ for all } x, y.$$

We denote the events $(X \leqslant x)$, $(Y \leqslant y)$ respectively by A, B. Then

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leqslant x,\ Y \leqslant y)\\ &= P(AB) \geqslant P(A)+P(B)-1\\ &= P(X \leqslant x)+P(Y \leqslant y)-1\\ &= F_X(x)+F_Y(y)-1. \end{aligned}$$

$\therefore\quad F_X(x)+F_Y(y)-1 \leqslant F_{X,Y}(x, y).$

Again $AB \subseteq A$, $AB \subseteq B$.

$\therefore\quad P(AB) \leqslant P(A)$, $P(AB) \leqslant P(B)$,

and hence, $\{P(AB)\}^2 \leqslant P(A)\,P(B)$.

$\therefore\quad P(X \leqslant x, Y \leqslant y) \leqslant \sqrt{P(X \leqslant x)\,.\,P(Y \leqslant y)}$

i.e., $F_{X,Y}(x, y) \leqslant \sqrt{F_X(x)\,F_Y(y)}$.

Ex. 49. *Let $f_X(x)$ and $f_Y(y)$ be two probability density functions with corresponding distribution functions $F_X(x)$ and $F_Y(y)$ respectively. Show that the function $f_{X,Y}(x, y)$ defined by*

$$f_{X,Y}(x, y)=f_X(x)\,f_Y(y)\,[1+k\{2F_X(x)-1\}\{2F_Y(y)-1\}]$$

where $k \in [-1, 1]$, is a possible density function of a two-dimensional distribution and show that the corresponding marginal density functions are $f_X(x)$, $f_Y(y)$ respectively.

We have $0 \leqslant F_X(x) \leqslant 1$ for all x and so $-1 \leqslant F_X(x)-1$.

$\therefore\quad 0 \leqslant F_X(x)$ and $-1 \leqslant F_X(x)-1$ give $-1 \leqslant 2F_X(x)-1$.

Also $F_X(x) \leqslant 1$ gives $2F_X(x) \leqslant 2$ and so $2F_X(x)-1 \leqslant 1$.

Thus $-1 \leqslant 2F_X(x)-1 \leqslant 1$. Similarly $-1 \leqslant 2F_Y(y)-1 \leqslant 1$.

Also it is given that $-1 \leqslant k \leqslant 1$.

$\therefore\quad -1 \leqslant k\,[2F_X(x)-1][2F_Y(y)-1] \leqslant 1$, since each of the factors lies in the interval $[-1, 1]$.

$\therefore\quad -k[2F_X(x)-1][2F_Y(y)-1] \leqslant 1$

or, $1+k[2F_X(x)-1][2F_Y(y)-1] \geqslant 0$

or, $f_X(x)\,f_Y(y)\,[1+k\{2F_X(x)-1\}\{2F_Y(y)-1\}] \geqslant 0$,

since $f_X(x) \geqslant 0$, $f_Y(y) \geqslant 0$ for all x, y

i.e., $f_{X,Y}(x, y) \geqslant 0$ for all x, y.

Again $\displaystyle\int_{-\infty}^{\infty} f_X(x)\,[2F(x)-1]\,dx=\int_0^1 (2u-1)\,du=0$

where $u=F_X(x)$, so that $du=F'_X(x)\,dx=f_X(x)\,dx$, and $u \to 1$ when $x \to +\infty$, $u \to 0$ when $x \to -\infty$.

Similarly, $\displaystyle\int_{-\infty}^{\infty} f_Y(y)\,[2F_Y(y)-1]\,dy=0.$

Now $\displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$

$$=\int_{-\infty}^{\infty} f_X(x)\,f_Y(y)\,[1+k\{2F_X(x)-1\}\{2F_Y(y)-1\}]\,dy$$

$$=f_X(x)\left[\int_{-\infty}^{\infty} f_Y(y)\,dy+k\{2F_X(x)-1\}\int_{-\infty}^{\infty}\{2F_Y(y)-1\}f_Y(y)\,dy\right]$$

$=f_X(x)$, since $\displaystyle\int_{-\infty}^{\infty} f_Y(y)\,dy=1$, $f_Y(y)$ being a probability density function and $\displaystyle\int_{-\infty}^{\infty}\{2F_Y(y)-1\}\,f_Y(y)\,dy=0.$

$$\therefore\quad \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy\right\}dx=\int_{-\infty}^{\infty} f_X(x)\,dx=1,$$

$f_X(x)$ being a probability density function.

Thus $f_{X,Y}(x, y) \geqslant 0$ for all x, y and $\displaystyle\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy\right\}dx=1.$

Hence, $f_{X,Y}(x, y)$ is a possible two-dimensional probability density function.

Also the relations $\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = f_X(x)$ and

$\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = f_Y(y)$ (which can be shown similarly) shows that $f_X(x)$ and $f_Y(y)$ are the corresponding marginal density functions.

Ex. 50. *Let F be the distribution function of a random variable. Examine whether the following functions are joint distribution functions :*

(*i*) $F(x)+F(y)$; (*ii*) $\max\{F(x), F(y)\}$.

(*i*) Here $\underset{x\to-\infty}{Lt}[F(x)+F(y)]$

$$=F(-\infty)+F(y)$$
$$=F(y) \quad [\because F(-\infty)=0].$$

But if $F(x)+F(y)$ be a joint distribution function, then

$$\underset{x\to-\infty}{Lt}[F(x)+F(y)]=0.$$

So $F(y)=0$ for all $y \in R$.

Again, F being a distribution function, we have $\underset{y\to\infty}{Lt} F(y)=1$.

But $F(y)=0$ for all $y \in R$

$$\Rightarrow \underset{y\to\infty}{Lt} F(y)=0.$$

So we arrive at a contradiction.

Hence, $F(x)+F(y)$ cannot be a joint distribution function.

(*ii*) Let $\varphi(x, y)=\max\{F(x), F(y)\}$.

Here we see that

$$\underset{x\to-\infty}{Lt}\varphi(x, y)=\underset{x\to-\infty}{Lt}\max\{F(x), F(y)\}$$
$$=\underset{x\to-\infty}{Lt} F(y),$$

[$\because$ here for any fixed y, $x < y$ and so $F(y) \geqslant F(x)$]
$$=F(y).$$

If $\varphi(x, y)$ be a joint distribution function, then
$$\underset{x\to-\infty}{Lt}\varphi(x, y)=0.$$

Hence $F(y)=0$ for all $y \in R$ and so we get
$$\underset{y\to\infty}{Lt} F(y)=0, \text{ that is, } 1=0 \text{ which is absurd.}$$

So $\max\{F(x), F(y)\}$ cannot be a joint distribution function.

Ex. 51. *Show that $F_1(x)+F_2(x)+F_1(y)+F_2(y)$ cannot be a two-dimensional distribution function, where F_1 and F_2 are two distribution functions in one-dimension.*

Let $F(x, y)=F_1(x)+F_2(x)+F_1(y)+F_2(y)$ and if possible let $F(x, y)$ be a two-dimensional distribution function.

Proceeding to the limit $y\to-\infty$,
$$0=F(x, -\infty)=F_1(x)+F_2(x)+F_1(-\infty)+F_2(-\infty)$$
$$=F_1(x)+F_2(x), \text{ since } F_1(-\infty)=F_2(-\infty)=0$$

Proceeding to the limit $x\to\infty$,
$$0=F_1(\infty)+F_2(\infty)=2, \text{ since } F_1(\infty)=F_2(\infty)=1$$

and we get a contradiction. Hence, $F(x, y)$ cannot be a two-dimensional distribution function.

Ex. 52. *The distribution function F of a two-dimensional random variable (X, Y) is given by*
$$F(x, y)=F_1(x)\,F_2(y)+F_3(x), \textit{ for all } x, y \in R.$$

Can the function $F_3(x)$ be arbitrary ? Are X, Y independent ?

We have $\underset{y\to-\infty}{Lt} F(x, y)=0.$

So $\underset{y\to-\infty}{Lt}[F_1(x)\,F_2(y)+F_3(x)]$ must exist finitely and its value is 0.

But $\underset{y\to-\infty}{Lt} F_3(x)=F_3(x)$, which is finite for given value of x.

So $\underset{y\to-\infty}{Lt} F_2(y)$ also exists finitely. Then we get

$$F_1(x)\underset{y\to-\infty}{Lt} F_2(y)+F_3(x)=0$$

or, $F_3(x)=a\,F_1(x)$ where $a=-\underset{y\to-\infty}{Lt} F_2(y)$ ... ... (6.10.4).

The relation $F_3(x)=a\,F_1(x)$ shows that $F_3(x)$ is not arbitrary.

Now the marginal distribution function F_X of X is given by

$$F_X(x)=\underset{y\to\infty}{Lt}\ F(x,y)=\underset{y\to\infty}{Lt}\ [\,F_1(x)\,F_2(y)+F_3(x)\,]$$

$$=\underset{y\to\infty}{Lt}\ F_1(x)\,F_2(y)+\underset{y\to\infty}{Lt}\ F_3(x)$$

(noting that the limits exist finitely for fixed x).

So $F_X(x)=b\,F_1(x)+F_3(x)$, where $b=\underset{y\to\infty}{Lt}\ F_2(y)$.

Hence, $F_X(x)=(b+a)\,F_1(x)$ by (6.10.4). ... . (6.10.5)

The marginal distribution function F_Y of Y is given by

$$F_Y(y)=\underset{x\to\infty}{Lt}\ F(x,y)=\underset{x\to\infty}{Lt}\ [\,F_1(x)\,F_2(y)+F_3(x)\,]$$

$$=F_2(y)\ \underset{x\to\infty}{Lt}\ F_1(x)+\underset{x\to\infty}{Lt}\ F_3(x).$$

[By (6.10.5), $\underset{x\to\infty}{Lt}\ F_1(x)$ exists finitely and then by (6.10.4) $\underset{x\to\infty}{Lt}\ F_3(x)$ exists finitely].

Thus we get $F_Y(y)=k_1F_2(y)+k_2$ where $k_1=\underset{x\to\infty}{Lt}\ F_1(x)$, $k_2=\underset{x\to\infty}{Lt}\ F_3(x)$.

Now by (6.10.5) and (6.10.4) we get (taking limit as $x\to\infty$)

$$1=(a+b)\,k_1,\ k_2=ak_1.$$

So $k_1\neq 0$, $a+b\neq 0$ and then we get

$$F_X(x)=\frac{1}{k_1}\,F_1(x), \qquad \text{... (6.10.6)}$$

and $F_Y(y)=k_1\,F_2(y)+ak_1$

or, $F_Y(y)=k_1\,[F_2(y)+a]$. ... (6.10.7)

Then
$$F(x,y)=F_1(x)\,F_2(y)+F_3(x)$$
$$=F_1(x)\,F_2(y)+aF_1(x)$$
$$=F_1(x)\,[\,F_2(y)+a\,]$$
$$=\frac{F_1(x)}{k_1}.\,k_1\,[\,F_2(y)+a\,]$$
$$=F_X(x)\,F_Y(y),\text{ by (6.10.6) and (6.10.7)}$$

Thus it is proved that $F(x,y)=F_X(x)\,F_Y(y)$ for all x, y.

So X, Y are independent.

Examples VI

1. Consider the random experiment of tossing two fair coins. Let X be the random variable taking values 1 or 0 according as the outcome is 'head' or 'tail' for the first coin and Y be the random variable taking values 1 or 0 according as the outcome is head or tail for the second coin. Find the distribution of (X, Y) and the marginal distributions of X and Y.

[*Hints* : The p.m.f. of (X, Y) is shown by the following table, along with the marginal p.m.f. of X and Y :

$X \backslash Y$	0	1	$P(X=x_i)$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$P(Y=y_j)$	$\frac{1}{2}$	$\frac{1}{2}$	1

2. Two dice are thrown. Let the random variable X denote the number shown by the first die and the random variable Y denote the larger of the two numbers. Find (*a*) the distribution of the two-dimensional random variable (X, Y) ; (*b*) the marginal distributions of X and Y ; (*c*) $P(Y=4 \mid X=3)$. Are the random variables X and Y independent ?

[*Hints* : (*a*), (*b*) The spectrum of (X, Y) is

$(x_i, y_j)=(i, j)$; $(i=1, 2, 3, 4, 5, 6$; $j=1, 2, 3, 4, 5, 6)$,

the event $(X=1, Y=1)$ contains only one outcome $(1, 1)$ and so $p_{11}=\frac{1}{36}$; the event $(X=2, Y=2)=\{(2, 1), (2, 2)\}$ and $p_{22}=\frac{2}{36}$; the event $(X=3, Y=1)$ is an impossible event and so $p_{31}=0$ and so on.

In tabular form, the p.m.f. of the two-dimensional random variable (X, Y) along with the marginal distributions of X and Y is given in the next page.

$X \backslash Y$	1	2	3	4	5	6	$P(X=x_i)$
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	0	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	0	0	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	0	0	0	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	0	0	0	0	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	0	0	0	0	0	$\frac{6}{36}$	$\frac{1}{6}$
$P(Y=y_j)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1

(c) $P(Y=4 \mid X=3)=\dfrac{P(X=3,\,Y=4)}{P(X=3)}=\dfrac{\frac{1}{36}}{\frac{1}{6}}=\frac{1}{6}.$

Now $P(X=1,\,Y=2)=\frac{1}{36}$ and $P(X=1)=\frac{1}{6}$, $P(Y=2)=\frac{2}{36}$;

$\therefore\ P(X=1,\,Y=2) \neq P(X=1)\,P(Y=2).$

Hence X and Y are not independent.]

3. The following table gives the p.m.f. of the two-dimensional random variable (X, Y):

$X \backslash Y$	1	2	3
1	$\frac{5}{27}$	$\frac{1}{27}$	$\frac{3}{27}$
2	$\frac{4}{27}$	$\frac{3}{27}$	$\frac{4}{27}$
3	$\frac{2}{27}$	$\frac{3}{27}$	$\frac{2}{27}$

Find (a) the marginal distributions of X and Y,

(b) the conditional distribution of X given $Y=2$ and of Y given $X=1$,

(c) $F(0, 1)$, $F(2, 3)$, $F(\frac{3}{5}, \frac{5}{2})$, $F(8, 10)$ where $F(x, y)$ is the distribution function of (X, Y).

Are X and Y independent?

[*Hints*: (a) The marginal distribution of X is:

$x_i=i\ (i=1, 2, 3)$ with $f_{xi}=P(X=x_i)$ where

$f_{x1}=\frac{9}{27},\ f_{x2}=\frac{11}{27},\ f_{x3}=\frac{7}{27}.$

The marginal distribution of Y is:

$y_j=j\ (j=1, 2, 3)$ with $f_{yj}=P(Y=y_j)$ where

$f_{y1}=\frac{11}{27},\ f_{y2}=\frac{7}{27},\ f_{y3}=\frac{9}{27}.$

(b) $f_x(1 \mid 2)=\frac{1}{7},\ f_x(2 \mid 2)=\frac{3}{7},\ f_x(3 \mid 2)=\frac{3}{7},$

$f_y(1 \mid 1)=\frac{5}{9},\ f_y(2 \mid 1)=\frac{1}{9},\ f_y(3 \mid 1)=\frac{1}{3}.$

(c) $F(0, 1)=\sum_{i\leq 0}\sum_{j\leq 1} f_{ij}=0.$

$F(2, 3)=\sum_{i\leq 2}\sum_{j\leq 3} f_{ij}=\sum_{i\leq 2}(f_{i1}+f_{i2}+f_{i3})$

$=(f_{11}+f_{12}+f_{13}+f_{21}+f_{22}+f_{23})=\frac{20}{27}.$

$F(4, 1)=\sum_{i\leq 4}\sum_{j\leq 1} f_{ij}=\sum_{i\leq 4} f_{i1}=(f_{11}+f_{21}+f_{31})=\frac{11}{27}.$

$F(\frac{3}{5}, \frac{5}{2})=0,\quad F(8, 10)=1.$

$f_{12}=\frac{1}{27},\ f_{x1}=\frac{9}{27},\ f_{y2}=\frac{11}{27}\quad \therefore\ f_{12} \neq f_{x1}f_{y2}$, so that X and Y are not independent.]

4. A two-dimensional random variable (X, Y) has the spectrum $(x_i, y_j)=(i, j)$, $(i=1, 2, 3;\ j=1, 2, 3)$ and the probability masses p_{ij} are given by

$p_{ij}=P(X=i,\ Y=j)=Kij$, where K is a constant.

Find (a) the value of K, (b) $P(1\leq X\leq 3,\ Y\leq 2)$, (c) $P(X\geq 2)$, (d) $P(Y\leq 2)$, (e) $P(X=2)$ (f) marginal p.m.f. of X and Y. Are X and Y independent?

Ex. VI

5. Show that the function $F(x, y)=\begin{cases} 0, & x+y<1 \\ 1, & x+y \geqslant 1 \end{cases}$

is not a distribution function.

[*Hints* : Use § 6.2, property 7.]

6. Let (X, Y) have a density function defined by $f(x, y)=\frac{1}{2}$ inside a square with corners at the points $(1, 0)$, $(0, 1)$, $(-1, 0)$ and $(0, -1)$ and equal to zero elsewhere. Find the marginal density functions of X and Y and the two conditional density functions.

[*Hints* : See Illustrative Ex. 7.

$$f_Y(y \mid x)=\begin{cases} \dfrac{1}{2(1-x)} & \text{if } x-1<y<1-x,\ 0<x<1 \\ \dfrac{1}{2(1+x)}, & \text{if } -1-x<y<1+x,\ -1<x<0. \end{cases}$$

$$f_X(x \mid y)=\begin{cases} \dfrac{1}{2(1-y)} & \text{for } y-1<x<1-y,\ 0<y<1 \\ \dfrac{1}{2(1+y)} & \text{for } -1-y<x<1+y,\ -1<y<0. \end{cases}$$]

7. If (X, Y) is uniform in the triangle $x \geqslant 0$, $y \geqslant 0$, $x+y \leqslant 2$, find (*a*) the density function of (X, Y), (*b*) the marginal density function of X, (*c*) the conditional density function of Y, given $X=x$.

[*Hints* : (*a*) $f(x, y)=\frac{1}{2}$, $x \geqslant 0, y \geqslant 0, x+y \leqslant 2$.

(*b*) $f_X(x)=\displaystyle\int_{-\infty}^{\infty} f(x, y)\, dy=\int_0^{2-x} \tfrac{1}{2}\, dy=\tfrac{1}{2}(2-x),\ 0 \leqslant x \leqslant 2.$

(*c*) $f_Y(y \mid x)=\dfrac{f(x, y)}{f_X(x)}=\dfrac{1}{2-x},\ 0 \leqslant y \leqslant 2-x,\ 0 \leqslant x<2.$]

8. A gun is aimed at a certain point (origin of co-ordinate system). The actual hit-point can be any point (X, Y) within a circle of radius R about the origin. Assume that the density function of (X, Y) is constant within the circle and let it be given by

$$f(x, y)=C, \text{ for } x^2+y^2<R^2$$
$$=0, \text{ elsewhere.}$$

Ex. VI

(*a*) Compute C,

(*b*) Show that $f_X(x)=\begin{cases} \dfrac{2\sqrt{R^2-x^2}}{\pi R^2}, & -R<x<R \\ 0 & \text{, elsewhere.} \end{cases}$

(*c*) Are the random variables X and Y independent ?

[*Hints* : (*a*) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy=1$

or, $\displaystyle\iint_{x^2+y^2<R^2} C\, dx\, dy=1.$

Taking $x=r\cos\theta$, $y=r\sin\theta$, since $\left|\dfrac{\partial(x, y)}{\partial(r, \theta)}\right|=r$, we have

$$C\int_0^R \int_0^{2\pi} r\, d\theta\, dr=1. \quad \therefore \quad C=\frac{1}{\pi R^2}.$$

(*b*) See Illustrative Ex. 9.

(*c*) We have $f_X(x) f_Y(y)=\dfrac{4\sqrt{R^2-x^2}\sqrt{R^2-y^2}}{\pi^2 R^4}$, which is in general not equal to the value of the joint density function $f(x, y)$. It then follows that X and Y are not independent.]

9. The density function of a two-dimensional random variable (X, Y) is $f(x, y)=\begin{cases} 1, & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1 \\ 0, & \text{elsewhere.} \end{cases}$

Find (*a*) $P(X+Y \leqslant 1)$, (*b*) $P(X^2+Y^2 \leqslant \frac{1}{3})$,
(*c*) $P(X \geqslant \frac{1}{3} \mid Y \leqslant \frac{1}{3})$.

[*Hints* : (*a*) $P(X+Y \leqslant 1)=P(Y \leqslant 1-X)=\displaystyle\int_0^1\left\{\int_0^{1-x} dy\right\}dx=\frac{1}{2}.$

(*b*) $P(X^2+Y^2 \leqslant \frac{1}{3})=P(Y \leqslant \sqrt{\tfrac{1}{3}-X^2})$

$$=\int_0^{\frac{1}{\sqrt{3}}}\left\{\int_0^{\sqrt{\frac{1}{3}-x^2}} dy\right\} dx=\frac{\pi}{12}.$$

(*c*) $P(X \geqslant \frac{1}{3} \mid Y \leqslant \frac{1}{3})=\dfrac{P(X \geqslant \frac{1}{3}, Y \leqslant \frac{1}{3})}{P(Y \leqslant \frac{1}{3})}=\dfrac{2}{3}.$]

10. Let X and Y be independent random variables each having the normal density $(0, \sigma)$. Find $P(X^2+Y^2 \leq 1)$.

[*Hints*: Use polar co-ordinates in evaluating

$$\iint\limits_{x^2+y^2 \leq 1} \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} dx\, dy.]$$

11. The density function of (X, Y) is

$$f(x, y)=ky(1-x-y),\ x>0,\ y>0,\ x+y<1$$
$$=0, \text{ elsewhere.}$$

Find (*a*) the value of k, (*b*) marginal density functions of X and Y, (*c*) $P(X < \frac{1}{2}, Y < \frac{1}{2})$.

[*Hints*: (*a*) $\int_0^1 \left[\int_0^{1-x} ky(1-x-y)dy\right] dx=1$ gives $k=24$.

(*b*) $f_X(x)=\begin{cases}\int_0^{1-x} 24y(1-x-y)dy=4(1-x)^3, & 0<x<1,\\ 0, & \text{elsewhere.}\end{cases}$

$f_Y(y)=\begin{cases}\int_0^{1-y} 24y(1-x-y)dx=12y(1-y)^2, & 0<y<1\\ 0, & \text{elsewhere.}\end{cases}$

(*c*) $P(0 \leq X \leq \frac{1}{2}, 0 \leq Y \leq \frac{1}{2})$

$$=24\int_0^{\frac{1}{2}}\int_0^{\frac{1}{2}} y(1-x-y)\, dx\, dy=\frac{5}{8}.]$$

12. The probability density function of a two-dimensional random variable (X, Y) is

$$f(x, y)=C(x+y),\ 0 \leq x, y \leq 2$$
$$=0, \text{ elsewhere.}$$

Find (*a*) the value of C, (*b*) $P(|X-Y| \leq 1)$

13. The density function of a two dimensional random variable (X, Y) is given by

$$f(x, y)=4xy,\ 0<x, y<1$$
$$=0 \quad, \text{ elsewhere.}$$

Find (*a*) the marginal density functions of X and Y,

(*b*) the marginal distribution functions of X and Y,

(*c*) $f_X(x \mid y)$, $f_Y(y \mid x)$,

(*d*) $P(X+Y>1)$.

14. The density function of a two-dimensional random variable (X, Y) is

$$f(x, y)=\begin{cases}3x^2-8xy+6y^2, & 0 \leq x \leq 1, 0 \leq y \leq 1\\ 0 & , \text{ elsewhere.}\end{cases}$$

Find $f_X(x \mid y)$, $f_Y(y \mid x)$. Show that X and Y are not independent.

15. Let the density function of (X, Y) be

$$f(x, y)=\begin{cases}\frac{4}{3}\left(xy+\frac{x^2}{2}\right), & 0<x<1, 0<y<2\\ 0 & , \text{ elsewhere.}\end{cases}$$

Find $P(Y<1 \mid X<\frac{1}{2})$.

[*Hints*: $f_X(x)=\frac{4x}{3}(2+x)$, $0<x<1$.

$$P(Y<1 \mid X<\tfrac{1}{2})=\frac{P(X<\frac{1}{2}, Y<1)}{P(X<\frac{1}{2})}$$

$$=\frac{\int_0^1\left\{\int_0^{\frac{1}{2}} f(x, y)\, dx\right\}dy}{\int_0^{\frac{1}{2}} f_X(x)\, dx}=\frac{2}{7}.]$$

16. A dart is thrown at random on a square target board having vertices (1, 0), (0, 1), (−1, 0) and (0, −1), the point at which the dart hits the board being (X, Y). Find the marginal density functions of X and Y and show that they are not independent.

[*Hints*: See Illustrative Ex. 7.]

17. Let X and Y be continuous random variables having joint density function $f(x, y)$ given by

$$f(x, y) = \lambda^2 e^{-\lambda y},\ 0 \leqslant x \leqslant y$$
$$= 0 \quad , \text{ elsewhere.}$$

Find the marginal density functions of X and Y. Find also the conditional density of Y given $X=x$ and that of X given $Y=y$.

[*Hints :* $f_X(x) = \lambda^2 \int_x^\infty e^{-\lambda y}\, dy = \lambda e^{-\lambda x},\ x > 0,$

$$f_Y(y) = \lambda^2 \int_0^y e^{-\lambda y}\, dx = \lambda^2 y\, e^{-\lambda y},\ y > 0.$$

The conditional density function of Y given $X=x$ is

$$f_Y(y \mid x) = \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)},\ 0 < x < y,$$

which is exponentially distributed (with change of origin).

Again, the conditional density function of X, given $Y=y$, is

$$f_X(x \mid y) = \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = \frac{1}{y},\ 0 < x < y,$$

which is uniform in the interval $(0, y)$.]

18. Given the joint density function of X and Y as

$$f(x, y) = \frac{2}{(1+x+y)^3},\ x > 0, y > 0.$$

Find (*a*) the joint distribution function $F(x, y)$,

(*b*) the marginal density function $f_X(x)$ of X,

(*c*) the conditional density function $f_Y(y \mid x)$,

given $X=x$.

[*Hints :* (*a*) $F(x, y) = 2 \int_0^x dx' \left[\int_0^y \frac{dy'}{(1+x'+y')^3} \right]$

$$= \int_0^x \left[\frac{1}{(1+x')^2} - \frac{1}{(1+x'+y)^2} \right] dx'$$

$$= 1 - \frac{1}{1+x} + \frac{1}{1+x+y} - \frac{1}{1+y},\ x > 0, y > 0.$$

(*b*) $f_X(x) = \int_0^\infty f(x, y)\, dy = \frac{1}{(1+x)^2},\ x > 0.$

(*c*) $f_Y(y \mid x) = \frac{2(1+x)^2}{(1+x+y)^3},\ y > 0, x > 0.$]

19. The density function of a two-demensional random variable is given by $f(x, y) = \begin{cases} Cx^2(8-y), & x < y < 2x,\ 0 \leqslant x \leqslant 2 \\ 0 & , \text{ elsewhere.} \end{cases}$

Find C, the marginal density functions of X and Y and the conditional density functions $f_X(x \mid y)$ and $f_Y(y \mid x)$.

[*Hints :* $C \int_0^2 x^2 \left[\int_x^{2x} (8-y)\, dy \right] dx = 1$ gives $C = \frac{5}{112}$.

$$f_X(x) = \tfrac{5}{112}(8x^3 - \tfrac{3}{2}x^4),\ 0 \leqslant x \leqslant 2,$$

$$f_Y(y) = \frac{5}{336}(8-y)\left(8 - \frac{y^3}{8}\right),\ 2 < y < 4,$$

$$= \tfrac{5}{384}\, y^3(8-y),\ 0 < y \leqslant 2,$$

$$f_X(x \mid y) = \frac{24x^2}{7y^3},\ \frac{y}{2} < x < y,\ 0 < y \leqslant 2,$$

$$= \frac{24x^2}{64 - y^3},\ \frac{y}{2} < x < y,\ 2 < y < 4.$$

$$f_Y(y \mid x) = \frac{6(8-y)}{x(16-3x)},\ x < y < 2x,\ 0 < x \leqslant 2.]$$

20. X and Y are independent random variables each uniformly distributed in (0, 1). Find the probability density functions of $X+Y$, $X-Y$, $\mid X-Y \mid$.

Find also the distribution function of $X+Y$.

[*Hints :* Let $U = X+Y$, $V = X-Y$. In terms of real variables $u = x+y$, $v = x-y$, so that $\frac{\partial(x, y)}{\partial(u, v)} = -\frac{1}{2}$. As the point (x, y) varies within the square in the xy-plane bounded by $x=0$, $y=0$, $x=1$, $y=1$, the point (u, v) varies within the square R in uv-plane bounded by the lines $u+v=0$, $u-v=0$, $u+v=2$, $u-v=2$. (See Fig. 6.10.42). Since X, Y are independent

$$f_{X,Y}(x, y) = 1,\ 0 < x, y < 1$$
$$= 0, \text{ elsewhere.}$$

$$\therefore \quad f_{U,V}(u, v) = 1 \times \frac{1}{|-2|} = \frac{1}{2}, \ (u, v) \in R$$
$$= 0, \text{ elsewhere.}$$

The marginal density functions of U and V are,

$$f_U(u) = \int_{-u}^{u} \tfrac{1}{2}\, dv = u, \ 0 < u < 1$$
$$= \int_{u-2}^{2-u} \tfrac{1}{2} dv = 2 - u, \ 1 < u < 2$$
$$= 0, \text{ elsewhere,}$$

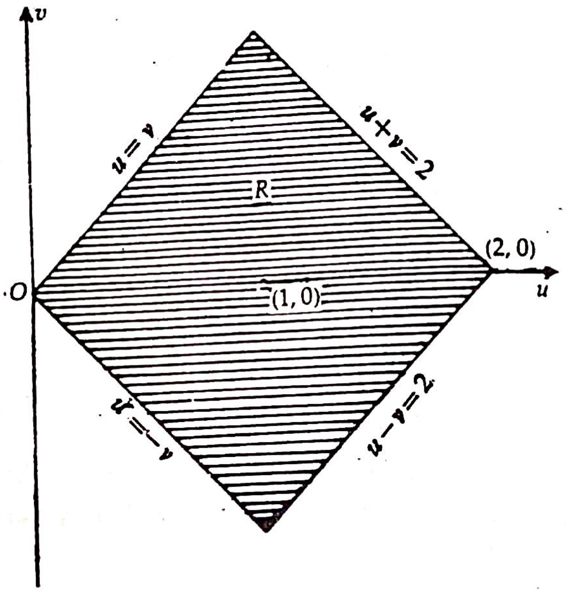


Fig. 6.10.42

and
$$f_V(v) = \int_{-v}^{2+v} \tfrac{1}{2} du = v + 1, \ -1 < v < 0$$
$$= \int_{v}^{2-v} \tfrac{1}{2} du = 1 - v, \ 0 < v < 1$$
$$= 0, \text{ elsewhere.}$$

The distribution function $F_U(u)$ of U is given by

$$F_U(u) = \int_{-\infty}^{u} f_U(u')\, du' = 0 \text{ for } u < 0$$
$$= \int_{0}^{u} u' du' = \frac{u^2}{2}, \text{ for } 0 < u < 1$$
$$= \int_{0}^{1} u' du' + \int_{1}^{u} (2 - u') du' = 1 - \tfrac{1}{2}(2-u)^2, \ 1 \leqslant u < 2$$
$$= 1 \text{ for } u \geqslant 2.$$

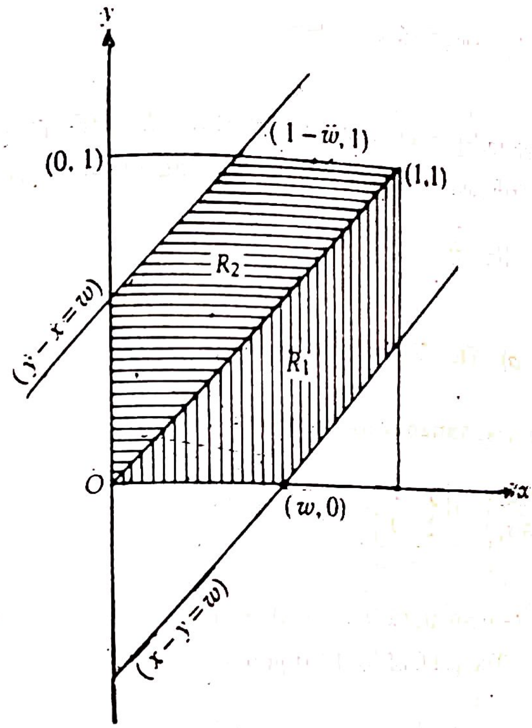


Fig. 6.10.43

Let $W = | X - Y |$. Then the distribution function of W is given by

$$F_W(w) = P(W \leqslant w) = P(\,| X - Y | \leqslant w), \ 0 \leqslant w \leqslant 1$$
$$= P(X - Y \leqslant w, \ X > Y) + P(Y - X \leqslant w, \ X < Y), \ 0 \leqslant w \leqslant 1$$
$$= P\{(X, Y) \in R_1\} + P\{(X, Y) \in R_2\},$$

where R_1 and R_2 are the regions shown in Fig. 6.10.43.

$$\therefore F_W(w)=\int_0^{1-w} dx \int_0^{x+w} dy+\int_{1-w}^{w} dx \int_0^{1} dy+\int_w^1 dx \int_{x-w}^{1} dy$$

$$=2w-w^2,\ 0 < w < 1.$$

$F_W(w)=1$ if $w > 1$, $F_W(w)=0$, elsewhere.

Hence, the density function of w is

$f_W(w)=2(1-w),\ 0 < w < 1.$]

21. If two independent random variables X and Y be each uniformly distributed in the interval (0, 1), find the distributions of (a) XY and (b) $\dfrac{X}{Y}$.

[*Hints*: (a) Put $U=XY, V=X$.

In terms of real variables $u=xy$, $v=x$.

$$\therefore \frac{\partial(u,v)}{\partial(x,y)}=\begin{vmatrix} y & x \\ 1 & 0 \end{vmatrix}=-x<0 \text{ for all } x \in (0,1).$$

As x, y vary from 0 to 1, u and v also both vary in the interval (0, 1). The p.d.f. of (U, V) is given by

$$f_{U,V}(u,v)=\frac{1}{x}=\frac{1}{v},\ 0<v<1,\ 0<\frac{u}{v}<1;$$

$$i.e.,\ \ 0<u<v<1.$$

Hence, the p.d.f. of U is given by

$$f_U(u)=\int_u^1 \frac{1}{v}\,dv=-\log u,\ 0<u<1.$$

(b) Put $U=\dfrac{X}{Y}, V=X.$

In real variables $u=\dfrac{x}{y}$, $v=x$, *i.e.*, $x=v$, $y=\dfrac{v}{u}$. As x, y vary from 0 to 1, u varies from 0 to ∞ and v varies in (0, 1).

$$\text{Then } \frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix} 0 & 1 \\ -\dfrac{v}{u^2} & \dfrac{1}{u} \end{vmatrix}=\frac{v}{u^2}>0 \text{ for all } u \in (0,\infty) \text{ and } v \in (0,1).$$

The p. d. f. of (U, V) is then given by

$$f_{U,V}(u,v)=\frac{v}{u^2},\ u>0,\ 0<v<u,\ v<1.$$

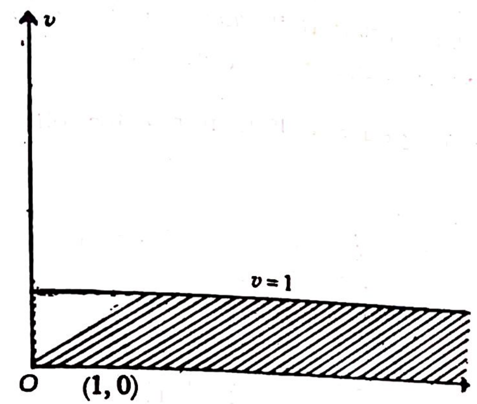


Fig. 6.10.44

Hence the p.d.f. of U is given by

$$f_U(u)=\int_0^u \frac{v}{u^2}\,dv=\tfrac{1}{2} \text{ if } 0<u<1,$$

$$=\int_0^1 \frac{v}{u^2}\,dv=\frac{1}{2u^2},\ u>1.]$$

22. The p.m.f. of the two-dimensional random variable (X, Y) is given by the following table :

Y \ X	-1	0	1	$P(Y=y_j)$
-2	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{5}{12}$
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{5}{12}$
2	$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{1}{6}$
$P(X=x_i)$	$\frac{5}{12}$	$\frac{1}{6}$	$\frac{5}{12}$	

Find the distribution of (U, V), where $U=|X|$, $V=Y^2$. Find also the marginal p.m.f. of U and V.

[*Hints* : The p.m.f. of (U, V) is given by the following table :

V \ U	0	1	$P(V=v_j)$
1	$\frac{1}{12}$	$\frac{1}{3}$	$\frac{5}{12}$
4	$\frac{1}{12}$	$\frac{1}{2}$	$\frac{7}{12}$
$P(U=u_i)$	$\frac{1}{6}$	$\frac{5}{6}$	1

Note that $P(U=0, V=1)=P(X=0, Y=1)=\frac{1}{12}$,

$$P(U=1, V=4)=P(X=1, Y=2)+P(X=1, Y=-2)$$
$$+P(X=-1, Y=2)+P(X=-1, Y=-2)$$
$$=\tfrac{1}{12}+\tfrac{1}{6}+\tfrac{1}{12}+\tfrac{1}{6}=\tfrac{1}{2} \text{ etc.},$$

The marginal p.m.f. of U is given by

$$f_{u\,i}=\begin{cases}\frac{1}{6}, & i=0\\ \frac{5}{6}, & i=1\end{cases}$$

and that of V is

$$f_{v\,j}=\begin{cases}\frac{5}{12}, & j=1\\ \frac{7}{12}, & j=4.\end{cases}\ ]$$

23. If X, Y are two independent random variables each following the distribution

$$f(x)=\tfrac{1}{2}e^{-\frac{x}{2}},\ x \geq 0,$$

find the distribution of $\frac{1}{2}(X-Y)$.

[*Hints* : Take $U=\frac{1}{2}(X-Y)$, $V=\frac{1}{2}(X+Y)$.

$$f_U(u)=\tfrac{1}{2}e^{-|u|},\ -\infty<u<\infty\ ]$$

24. Two independent random variables X and Y have the density functions

$$f_X(x)=3e^{-3x},\ x \geq 0$$

and

$$f_Y(y)=2e^{-2y},\ y \geq 0$$

respectively. Find the joint density functions of U, V, where $U=X+2Y$, $V=3X+4Y$. What are the marginal p.d.f. of U and V? Are U and V independent ?

[*Hints* : In real variables $u=x+2y$, $v=3x+4y$.

$\therefore\ x=v-2u,\ y=\frac{3}{2}u-\frac{1}{2}v.$

Since X, Y are independent, the joint density function is

$$f_{X,Y}(x, y)=6e^{-(3x+2y)},\ x \geq 0, y \geq 0.$$

$$J=\begin{vmatrix}-2 & 1\\ \frac{3}{2} & -\frac{1}{2}\end{vmatrix}=-\tfrac{1}{2}<0 \text{ for all } x \geq 0, y \geq 0.$$

Now $x \geq 0$, $y \geq 0$ implies $v \geq 2u$, $3u \geq v$,

i.e., $2u \leq v \leq 3u$.

Hence, the joint density function of (U, V) is

$$f_{U,V}(u, v)=6\,.\,e^{3u-2v},\ 2u < v < 3u$$
$$=0,\ \text{elsewhere.}$$

$\therefore$ the marginal density function of U is

$$f_U(u)=6\int_{2u}^{3u} e^{3u-2v}\,dv$$
$$=3(e^{-u}-e^{-3u}),\ u > 0,$$

and that of V is

$$f_V(v)=6\int_{\frac{v}{3}}^{\frac{v}{2}} e^{3u-2v}du=2(e^{-\frac{v}{2}}-e^{-v}),\ v > 0$$

$$\left(\because\ 2u < v < 3u \text{ implies } \frac{v}{3} < u < \frac{v}{2}\right).$$

Since $f_{U,V}(u, v) \neq f_U(u)\,f_V(v)$, U and V are not independent.]

25. If X and Y are independent binomial (n_1, p) and (n_2, p) variates respectively, show that $U=X+Y$ is a binomial (n_1+n_2, p) variate.

[*Hints*: Let $U=X+Y$.

The spectrum of U is the set $\{0, 1, 2, \ldots\ldots, n_1+n_2\}$.

Here $P(X=i)=\binom{n_1}{i}p^i(1-p)^{n_1-i}$ for $i=0, 1, 2, \ldots\ldots n_1$

and $P(Y=j)=\binom{n_2}{j}p^j(1-p)^{n_2-j}$ for $j=0, 1, 2, \ldots\ldots, n_2$.

Then for any given non-negative integer k,

$$P(U=k)=\sum_{i+j=k} P(X=i,\ Y=j)$$
$$=\sum_{i+j=k} P(X=i)\,P(Y=j), \text{ since } X,\ Y \text{ are independent.}$$

$$P(U=k)=\sum_{i=0}^{k}\binom{n_1}{i}\binom{n_2}{k-i}p^i(1-p)^{n_1-i}p^{k-i}(1-p)^{n_2-k+i}$$

$$=\sum_{i=0}^{k}\binom{n_1}{i}\binom{n_2}{k-i}p^k(1-p)^{n_1+n_2-k}.$$

Now from the identity $(1+x)^{n_1+n_2}=(1+x)^{n_1}(1+x)^{n_2}$, equating the coefficients of x^k from both sides we get

$$\binom{n_1+n_2}{k}=\sum_{i=0}^{k}\binom{n_1}{i}\binom{n_2}{k-i}.$$

$$\therefore\ P(U=k)=p^k(1-p)^{n_1+n_2-k}\sum_{i=0}^{k}\binom{n_1}{i}\binom{n_2}{k-i}$$

$$=\binom{n_1+n_2}{k}p^k(1+p)^{n_1+n_2-k},$$

$$k=0, 1, 2, \ldots\ldots, n_1+n_2,$$

which shows that $U=X+Y$ is a binomial (n_1+n_2, p) variate.]

26. Two points are independently chosen at random in the interval $(0, 1)$. Show that the probability that the distance between them is less than a fixed number c $(0 < c < 1)$ is $c\,(2-c)$.

[*Hints*: Let the random variables X and Y denote the points chosen in the interval $(0, 1)$. Then X and Y are both uniformly distributed in $(0, 1)$. The required event is $|X-Y| < c$.]

27. The joint probability density function of the random variables X, Y is

$$f_{X,Y}(x, y)=k(3x+y) \text{ when } 1 \leqslant x \leqslant 3,\ 0 \leqslant y \leqslant 2$$
$$=0,\ \text{elsewhere.}$$

Find (ii) $P(X+Y<2)$, (ii) the marginal distributions of X and Y. Investigate whether X and Y are independent.

[C. H. Math. '94]

[*Hints* : (i) from $k\int_1^3\left\{\int_0^2 (3x+y)\,dy\right\}dx=1$, we get $k=\frac{1}{28}$.

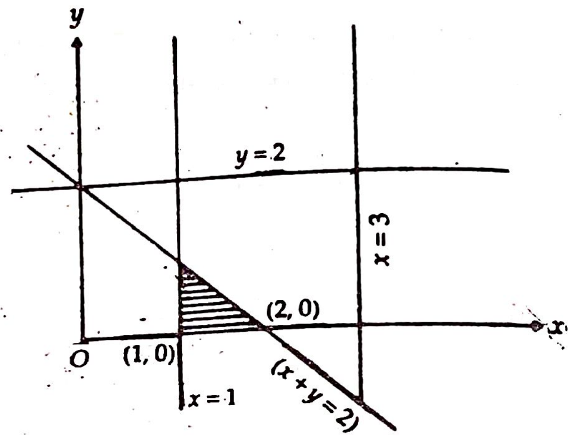


Fig. 6.10.44

$P(X+Y<2)=P\{(X,Y)\in R\}$ where R is the region shaded in Fig. 6.10.44,

$$=\iint_R \frac{1}{28}(3x+y)\,dx\,dy$$

$$=\frac{1}{28}\int_1^2\left\{\int_0^{2-x}(3x+y)\,dy\right\}dx$$

$$=\frac{13}{168}.$$

(ii) The marginal density function of X is given by

$$f_X(x)=\frac{1}{28}\int_0^2 (3x+y)\,dy, \text{ if } 1\leqslant x\leqslant 3$$

$$=0, \text{ elsewhere.}$$

i.e., $f_X(x)=\frac{3x+1}{14}$, if $1\leqslant x\leqslant 3$

$=0$, elsewhere.

Similarly, the marginal density function of Y is

$$f_Y(y)=\frac{y+6}{14}, \text{ if } 0\leqslant y\leqslant 2$$

$$=0, \text{ elsewhere.}$$

Since $f_{X,Y}(x,y)\neq f_X(x)\,f_Y(y)$, X and Y are not independent.]

28. Let X be a random variable having Poisson distribution with parameter λ and the conditional distribution of Y given $X=i$ be given by

$$f_{j|i}=\binom{i}{j}p^j q^{i-j}, \text{ for } 0\leqslant j\leqslant i,\ i\neq 0,\ p+q=1$$

$$=1, \text{ for } i=0.$$

Find the marginal distribution of Y.

[*Hints* : By definition $f_{ij}=f_{j|i}\cdot f_{xi}$, $f_{xi}=e^{-\lambda}\frac{\lambda^i}{i!}$, $i=0,1,2,\ldots$

$$\therefore\ f_{ij}=e^{-\lambda}\frac{\lambda^i}{i!}\binom{i}{j}p^j q^{i-j},\ 0\leqslant j\leqslant i,\ i=1,2,\ldots\ldots$$

$$=e^{-\lambda}\qquad,\ i=0.$$

The marginal distribution of Y is given by

$$f_{yj}=\sum_i f_{ij}=e^{-\lambda}+\sum_{i=j}^{\infty}e^{-\lambda}\frac{\lambda^i}{i!}\binom{i}{j}p^j q^{i-j}$$

$$=\frac{(\lambda p)^j}{j!}e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda q)^k}{k!},\ k=i-j$$

$$=\frac{e^{-\lambda}(\lambda p)^j}{j!}e^{\lambda q}$$

$$=\frac{e^{-\lambda(1-q)}(\lambda p)^j}{j!}$$

$$=\frac{e^{-\lambda p}(\lambda p)^j}{j!}.$$

Hence, the marginal distribution of Y is a Poisson distribution with parameter λp.]

29. The base x and altitude y of a triangle are segments taken at random 'of two straight lines of length a and b respectively. What is the probability that the area of the triangle is less then $\frac{ab}{4}$?

[*Hints* : Let X and Y denote the length and height of the triangle. Then X and Y are independent and uniformly distributed in the intervals $(0, a)$ and $(0, b)$ respectively.

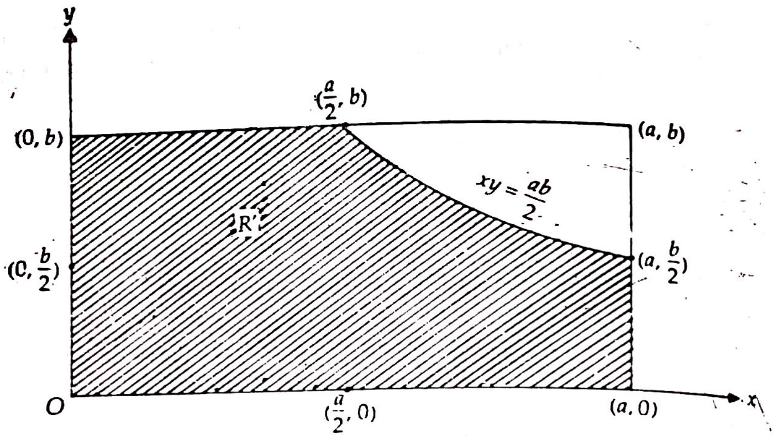


Fig. 6.10.45

$$\text{Required probability} = P\left(\tfrac{1}{2}XY < \frac{ab}{4}\right)$$

$$= P\left(XY < \frac{ab}{2}\right)$$

$$= \frac{R'}{R}$$

where R' is the shaded region in Fig. 6.10.45 and $R = ab$

$$= \frac{\frac{ab}{2} + \int_{\frac{a}{2}}^{a} \frac{ab}{2x}\,dx}{ab}$$

$$= \frac{1 + \log 2}{2}.\ ]$$

30. Two points P and Q are taken at random on a segment of length a. What is the probability that the distance between P and Q is less than b $(< a)$.

[*Hints* : Let the random variables X, Y denote the distances AP and AQ respectively, A being one end of the segment of length a.

We are to find the probability $P(|Y - X| < b)$.]

31. Two points are selected at random on a line of length a. What is the probability that none of the three sections into which the line thus divided is less than $\frac{1}{4}a$?

[*Hints* : Let P, Q be the two points chosen at random from the segment AB and let the random variables X and Y denote the lengths AP and BQ.

Fig. 6.10.46

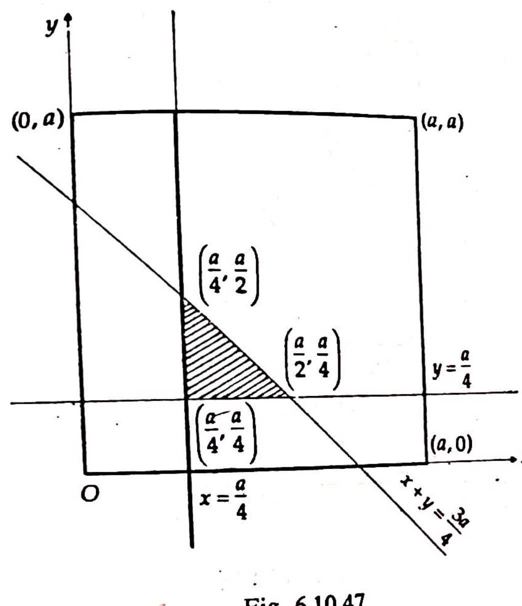


Fig. 6.10.47

Then $PQ=a-X-Y$, where $AB=a$. X and Y are independent, both uniformly distributed in the interval $(0, a)$. Now none of the three sections AP, PQ and QB is less than $\frac{1}{4}a$

$$\text{if } X \geqslant \frac{a}{4},\ Y \geqslant \frac{a}{4} \text{ and } a-(X+Y) \geqslant \frac{a}{4},$$

$$i.e., \text{ if } X \geqslant \frac{a}{4},\ Y \geqslant \frac{a}{4},\ X+Y \leqslant \frac{3a}{4}.$$

$$\text{Required probability}=\frac{R'}{R},$$

where R' is the shaded region in Fig. 6.10.47 and R is the square of area a^2.

$$\therefore \text{ required probability}=\frac{\frac{1}{2}\left(\frac{a}{4}\right)^2}{a^2}=\frac{1}{32}. \text{]}$$

32. Two points are chosen at random in the interval $(-1, 1)$. Find the probability that the three parts into which the interval is divided can form the sides of a triangle.

[*Hints* :

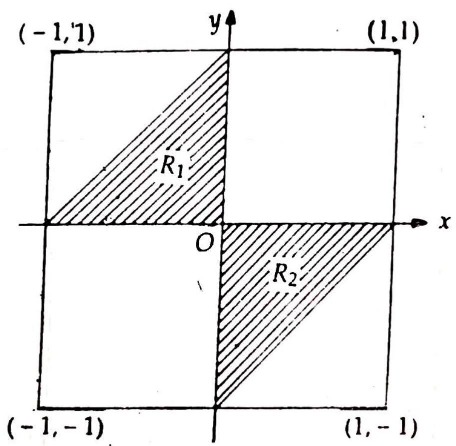


Fig. 6.10.48

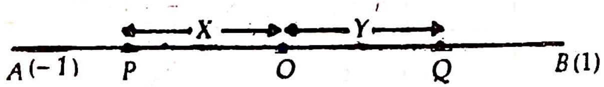


Fig. 6.10.49

Let the random variables X and Y denote the signed distances OP and OQ, P and Q being the points chosen at random from the interval $(-1, 1)$, O being the origin. Let $Y>X$. Then the three parts form a triangle if

$$AP+PQ>QB \quad \text{or} \quad X+1+Y-X>1-Y, \ i.e., \ Y>0,$$

$$PQ+QB>AP \quad \text{or} \quad Y-X+1-Y>X+1, \ i.e., \ X<0$$

and $$AP+QB>PQ \quad \text{or} \quad X+1+1-Y>Y-X, \ i.e., \ X-Y+1>0.$$

If $X>Y$, we get the corresponding inequalities as

$$X>0,\ Y<0 \text{ and } Y-X+1>0.$$

Required probability$=\frac{R_1+R_2}{4}=\frac{\frac{1}{2}+\frac{1}{2}}{4}=\frac{1}{4}$, R_1 and R_2 are the shaded regions in Fig. 6.10.48.]

33. If $F(x)$ is distribution function then prove that

(*i*) $F(x, y)=F(x)\,F(y)$ is a joint distribution function.

(*ii*) $F(x, y)=\min\,[F(x), F(y)]$ is a joint distribution function.

Answers

4. (*a*) $K=\frac{1}{36}$, (*b*) $\frac{5}{6}$, (*d*) $\frac{1}{2}$, (*e*) $f_{xi}=\frac{i}{6}$, $i=1, 2, 3$, $f_{yj}=\frac{j}{6}$, $j=1, 2, 3$. X and Y are independent.

10. $1-e^{-\frac{1}{2}\sigma^2}$ **12.** (*i*) $C=\frac{1}{8}$, (*ii*) $\frac{3}{4}$.

13. (*a*) $f_X(x)=2x,\ f_Y(y)=2y,\ 0<x, y<1.$

(*b*) $$F_X(x)=\begin{cases} x^2, & 0<x<1 \\ 1, & x \geqslant 1 \\ 0, & \text{elsewhere} \end{cases}, \quad F_Y(y)=\begin{cases} y^2, & 0<y<1 \\ 1, & y \geqslant 1 \\ 0, & \text{elsewhere.} \end{cases}$$

(*c*) $f_X(x \mid y)=2x,\ f_Y(y \mid x)=2y,\ 0<x<1,\ 0<y<1.$

(*d*) $P(X+Y>1)=P(1-X<Y<1)$

$$=\int_0^1 \left[\int_{1-x}^{1} 4xy\,dx\right]dx=\frac{5}{6}.$$

14. $\dfrac{3x^2-8xy+6y^2}{6y^2-4y+1}$, $\dfrac{3x^2-8xy+6y^2}{3x^2-4x+2}$, $0<x, y<1.$

30. $\dfrac{b}{a}\left(2-\dfrac{b}{a}\right).$

Here $f(x)$ has no local maximum and so X has no mode.

Now $F(\frac{5}{4})=\frac{5}{4}-\frac{1}{2}=\frac{3}{4}$, since $1<\frac{5}{4}<\frac{3}{2}$

and $F(\frac{1}{4})=\frac{1}{4}$, since $0<\frac{1}{4}<\frac{1}{2}$.

Hence $\zeta_{\frac{3}{4}}=\frac{5}{4}$ and $\zeta_{\frac{1}{4}}=\frac{1}{4}$.

Here the semi-interquartile range is

$$\tfrac{1}{2}(\zeta_{\frac{3}{4}}-\zeta_{\frac{1}{4}})=\tfrac{1}{2}(\tfrac{5}{4}-\tfrac{1}{4})=\tfrac{1}{2}.$$

7.8. Illustrative Examples :

Ex. 1. *A point P is chosen at random on a line segment AB of length $2l$. Find the expected values of (i) $AP \cdot PB$, (ii) $|AP-PB|$, (iii) $\max\{AP, PB\}$.*

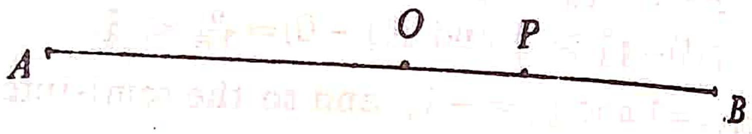


Fig. 7.8.1

Let O be the middle point of AB and X be the random variable denoting the length of OP prefixed with proper sign. Then X has uniform distribution in $(-l, l)$. So if $f(x)$ be the probability density function of X, then

$$f(x)=\frac{1}{2l} \text{ if } -l<x<l$$

$$=0, \text{ elsewhere.}$$

(i) $E(AP \cdot PB)=E\{(l+X)(l-X)\}=E(l^2-X^2)$

$$=\int_{-l}^{l}(l^2-x^2)\cdot\frac{1}{2l}\,dx$$

$$=\frac{1}{2l}\left(2l^3-\frac{2l^3}{3}\right)=\tfrac{2}{3}l^2.$$

(ii) $E(|AP-PB|)=E(|l+X-l+X|)=E(|2X|)$

$$=\int_{-l}^{l}|2x|\frac{1}{2l}\,dx=\int_{-l}^{l}\frac{|x|}{l}\,dx$$

$$=\frac{2}{l}\int_{0}^{l}|x|\,dx=\frac{2}{l}\int_{0}^{l}x\,dx=l.$$

(iii) $E[\max\{AP, PB\}]$

$$=E[\max\{l+X, l-X\}]$$

$$=\int_{-l}^{l} \max\{l+x, l-x\} \frac{1}{2l}\, dx$$

$$=\frac{1}{2l}\int_{-l}^{0} \max\{l+x, l-x\}\, dx+\frac{1}{2l}\int_{0}^{l} \max\{l+x, l-x\}\, dx.$$

Now, $\max\{l+x, l-x\}=l+x$ if $x \geqslant 0$

and $\max\{l+x, l-x\}=l-x$ if $x \leqslant 0$.

So, $E[\max\{l+X, l-X\}]$

$$=\frac{1}{2l}\int_{-l}^{0} (l-x)\, dx+\frac{1}{2l}\int_{0}^{l} (l+x)\, dx$$

$$=\frac{1}{2l}\left(l^2+\frac{l^2}{2}\right)+\frac{1}{2l}\left(l^2+\frac{l^2}{2}\right)=\frac{3l}{2}.$$

Ex. 2. *Find $E(X)$ for the following density function :*

$$f(x)=\frac{4x}{5} \quad \text{when } 0<x\leqslant 1$$

$$=\tfrac{2}{5}(3-x) \quad \text{when } 1<x\leqslant 2$$

$$=0, \quad \text{elsewhere.}$$

[*C. H. (Econ.) '91*]

$$E(X)=\int_0^2 x f(x)\, dx$$

$$=\int_0^1 x\left(\frac{4x}{5}\right) dx+\int_1^2 x\cdot\tfrac{2}{5}(3-x)\, dx$$

$$=\tfrac{4}{5}\cdot\tfrac{1}{3}+\tfrac{2}{5}\left(\tfrac{10}{3}-\tfrac{7}{6}\right)=\tfrac{17}{15}.$$

Ex. 3. *If the probability density function of a random variable X is given by $f(x)=C\,e^{-(x^2+2x+3)}$, $-\infty<x<\infty$, find the value of C, the expectation and variance of the distribution.* 2013

[*C. H. (Math.) '89*]

We have $C\int_{-\infty}^{\infty} e^{-(x^2+2x+3)}\, dx=1$

or, $C\int_{-\infty}^{\infty} e^{-(x+1)^2} e^{-2}\, dx=1$

or, $Ce^{-2}\int_{-\infty}^{\infty} e^{-z^2}\, dz=1$, $(x+1=z)$ $\Rightarrow 2Ce^{-2}\int_0^{\infty} e^{-z^2} d$

or, $Ce^{-2}\Gamma(\tfrac{1}{2})=1$.

$\therefore\ C=\dfrac{e^2}{\sqrt{\pi}}$.

-is an even ...tion

$z^2=t \Rightarrow 2z\,d$

$\Rightarrow 2Ce^{-2}\int t^{-\frac{1}{2}}$

$$E(X)=\frac{e^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} x\, e^{-(x^2+2x+3)}\, dx$$

$$=\frac{e^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-2}x\, e^{-(x+1)^2}\, dx$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} x\, e^{-(x+1)^2}\, dx$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z-1)\, e^{-z^2}\, dz,\ (x+1=z).$$

Now, $\int_{-\infty}^{\infty} z\, e^{-z^2}\, dz$ is convergent and its value is zero.

$$\therefore\ E(X)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (-e^{-z^2})\, dz$$

$$=\frac{1}{\sqrt{\pi}}(-\sqrt{\pi})=-1.$$

Again, $E(X^2)=\frac{e^2}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^2\, e^{-(x^2+2x+3)}\, dx$

$$=\frac{e^2e^{-2}}{\sqrt{\pi}}\int_{-\infty}^{\infty} x^2 e^{-(x+1)^2}\, dx$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z-1)^2 e^{-z^2}\, dz$$

$$=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} (z^2-2z+1)e^{-z^2}\, dz.$$

Now, $\int_{-\infty}^{\infty} z^2e^{-z^2}dz=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\int_{B_1}^{B_2} z^2e^{-z^2}\, dz$

$$=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\left(\int_{B_1}^{0} z^2e^{-z^2}dz+\int_{0}^{B_2} z^2 e^{-z^2}\, dz\right)$$

$$=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\left(\int_{B_1^2}^{0} \frac{u\, e^{-u}}{-2\sqrt{u}}\, du+\int_{0}^{B_2^2} \frac{u\, e^{-u}du}{2\sqrt{u}}\right)$$

$$=\underset{\substack{B_2\to\infty\\ B_1\to-\infty}}{Lt}\left(\frac{1}{2}\int_{0}^{B_1^2} \sqrt{u}\, e^{-u}\, du+\frac{1}{2}\int_{0}^{B_2^2}\sqrt{u}\, e^{-u}du\right)$$

$$=\int_0^{\infty} u^{\frac{3}{2}-1}\, e^{-u}\, du=\Gamma(\tfrac{3}{2})=\tfrac{1}{2}\sqrt{\pi}.$$

$z^2e^{-z^2}dz$ [∵ $z^2e^{-z^2}$ is an even function]

2013

$z^2=t$

$2z\,dz=dt$

$t^{\frac{3}{2}-1}e^{-t}dt$

$\Gamma\frac{3}{2}$

$=\frac{1}{2}\Gamma\frac{1}{2}+1$

$=\frac{1}{2}\Gamma\frac{1}{2}$

$=\frac{\sqrt{\pi}}{2}$

Thus we see that $E(X_i)=\dfrac{1}{p}$ and this is true for every i. Hence, the required expectation of shell consumption is

$$E(X_1+X_2+\cdots+X_{10})$$
$$=E(X_1)+E(X_2)+\cdots+E(X_{10}) \qquad \text{(see chapter VIII)}$$
$$=10\ E(X_1)=\frac{10}{p}.$$

[It will be proved in the next chapter that if $E(X_1), E(X_2),\ldots, E(X_n)$ exist, then
$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n).\ ]$$

Ex. 8. *If t is a positive real number and the probability mass function of a discrete random variable X is given by*

$$f(x)=e^{-t}\,(1-e^{-t})^{x-1} \quad \textit{for} \quad x=1,2,3,\ldots$$
$$=0, \qquad \textit{elsewhere},$$

then find the mean and variance of X.

$$E(X)=\sum_{x=1}^{\infty} e^{-t}\,(1-e^{-t})^{x-1}.\,x$$
$$=e^{-t}\sum_{x=1}^{\infty} x\,(1-e^{-t})^{x-1}$$
$$=e^{-t}\,\{1+2\,(1-e^{-t})+3\,(1-e^{-t})^2+\cdots\}$$
$$=e^{-t}\,\{1-(1-e^{-t})\}^{-2}, \quad \text{since here } 0<1-e^{-t}<1.$$
$$=e^{-t}\,e^{2t}=e^t.$$

So the required mean is e^t.

Again, $E\{X(X-1)\}$

$$=\sum_{x=1}^{\infty} x\,(x-1)\,e^{-t}\,(1-e^{-t})^{x-1}$$
$$=e^{-t}\sum_{x=2}^{\infty} x(x-1)\,(1-e^{-t})^{x-1}$$
$$=e^{-t}\,\{2.\,1.\,(1-e^{-t})+3.2\,(1-e^{-t})^2+4.3\,(1-e^{-t})^3+\cdots\}$$
$$=e^{-t}\,(1-e^{-t})\,\{1.2+3.2\,(1-e^{-t})+3.4\,(1-e^{-t})^2+\cdots\},$$

where the infinite series within the second bracket is absolutely convergent for all $t>0$.

Now we know that, for $|x| < 1$,

$$(1-x)^{-1}=1+x+x^2+x^3+\cdots\cdots \quad (7.8.1)$$

The right hand side of (7.8.1) being a power series in x, we can differentiate both sides of (7.8.1) any number of times for $|x|<1$.

Differentiating twice with respect to x both sides of (7.8.1), we get

$$\frac{2}{(1-x)^3}=2.1+3.2.\ x+4.3.\ x^2+\cdots$$

Now, $0 < 1-e^{-t} < 1$, since we have $t > 0$.

So, $1.2+2.3\ (1-e^{-t})+3.4.\ (1-e^{-t})^2+\cdots\cdots$

$$=\frac{2}{\{1-(1-e^{-t})\}^3}=2e^{3t}.$$

Hence, we get $E[X(X-1)]=e^{-t}\ (1-e^{-t})\ 2e^{3t}$

$$=2e^{2t}\ (1-e^{-t}).$$

So the required variance is $E\ [X(X-1)]-m(m-1)$

$$=2e^{2t}\ (1-e^{-t})-e^{t}\ (e^{t}-1) \qquad [\because\ m=E(X)=e^t]$$

$$=e^{2t}-e^{t}.$$

Ex. 9. *A special die with $n+1$ faces is marked in its faces the numbers $0, \frac{1}{n}, \frac{2}{n}, \ldots\ldots, \frac{n-1}{n}, \frac{n}{n}$. The die is unbiased. Let X be the random variable denoting the number on the uppermost face. Find (a) $E(X)$, (b) $Var(X)$ and (c) coefficient of skewness of the distribution of X.*

Here $P\left(X=\frac{i}{n}\right)=\frac{1}{n+1}$ for $i=0, 1, 2, \cdots\cdots n$.

$$E(X)=\sum_{i=0}^{n} \frac{i}{n}\cdot \frac{1}{n+1}=\frac{1}{n(n+1)}\ (1+2+\cdots+n)=\tfrac{1}{2}.$$

$$\text{Var}\ (X)=E(X^2)-\{E(X)\}^2$$

$$=\frac{1}{n+1}\sum_{i=0}^{n}\frac{i^2}{n^2}-\tfrac{1}{4}.$$

$$=\frac{1}{n^2(n+1)}\cdot\frac{n(n+1)(2n+1)}{6}-\tfrac{1}{4}$$

$$=\frac{2n+1}{6n}-\tfrac{1}{4}=\frac{n+2}{12n}.$$

(c) Coefficient of skewness is given by

$$\gamma_1 = \frac{\mu_3}{\sigma^3}.$$

Now, $\mu_3 = E\{(X-\frac{1}{2})^3\}$

$$= E(X^3) - \tfrac{3}{2}E(X^2) + \tfrac{3}{4}E(X) - \tfrac{1}{8}$$

$$= \frac{1}{n+1}\sum_{i=0}^{n}\left(\frac{i}{n}\right)^3 - \frac{3}{2}\,\frac{2n+1}{6n} + \frac{3}{8} - \frac{1}{8}$$

$$= \frac{1}{n^3(n+1)}\left\{\frac{n(n+1)}{2}\right\}^2 - \frac{2n+1}{4n} + \frac{1}{4}$$

$$= \frac{n+1}{4n} - \frac{2n+1}{4n} + \frac{1}{4}$$

$$= \frac{n+1-2n-1+n}{4n} = 0.$$

Hence, $\gamma_1 = 0$.

2011

Ex. 10. *Find the mean, variance and the coefficient of skewness of the continuous distribution with probability density function given by*

$$f(x) = 1 - |1-x|, \quad 0 < x < 2$$
$$= 0 \qquad , \textit{elsewhere}.$$

If X be the corresponding random variable, then the mean is

$$E(X) = \int_0^2 (1 - |1-x|)\, x\, dx$$

$$= \int_0^1 (1 - |1-x|)\, x\, dx + \int_1^2 (1-|1-x|)\, x\, dx$$

$$= \int_0^1 \{1-(1-x)\}\, x\, dx + \int_1^2 \{1-(x-1)\}\, x\, dx$$

$$= \int_0^1 x^2\, dx + \int_1^2 (2x - x^2)\, dx$$

$$= \tfrac{1}{3} + \tfrac{2}{3} = 1.$$

Now $E(X^2)=\int_0^2 x^2 (1-|1-x|)\,dx$

$$=\int_0^1 x^2(1-|1-x|)\,dx+\int_1^2 x^2(1-|1-x|)\,dx$$

$$=\int_0^1 x^2(1-1+x)\,dx+\int_1^2 x^2(1-x+1)\,dx$$

$$=\int_0^1 x^3\,dx+\int_1^2 (2x^2-x^3)\,dx$$

$$=\tfrac{1}{4}+(\tfrac{16}{3}-\tfrac{16}{4})-(\tfrac{2}{3}-\tfrac{1}{4})$$

$$=\tfrac{1}{4}+\tfrac{16}{12}-\tfrac{5}{12}=\tfrac{7}{6}.$$

So, var $(X)=E(X^2)-\{E(X)\}^2=\frac{7}{6}-1=\frac{1}{6}.$

$$\mu_3'=E\{(X-1)^3\}$$

$$=E(X^3)-3E(X^2)+3E(X)-1$$

$$=E(X^3)-3.\tfrac{7}{6}+3-1$$

$$=E(X^3)-\tfrac{7}{2}+2$$

$$=\int_0^1 x^3(1-1+x)\,dx+\int_1^2 x^3(1-x+1)\,dx-\tfrac{3}{2}$$

$$=\int_0^1 x^4\,dx+\int_1^2 (2x^3-x^4)\,dx-\tfrac{3}{2}$$

$$=\tfrac{1}{5}+\tfrac{8}{5}-\tfrac{3}{10}-\tfrac{3}{2}=0.$$

So, coefficient of skewness is

$$\gamma_1=\frac{\mu_3}{\sigma^3}=0.$$

Ex. 11. *If the roots of the quadratic equation* $t^2-at+b=0$ *are real and* b *is positive but otherwise unknown, then find the mean values of the roots.*

The roots of the equation $t^2-at+b=0$ are

$$\frac{a+\sqrt{a^2-4b}}{2},\quad \frac{a-\sqrt{a^2-4b}}{2},$$

where $a^2-4b>0.$

If X_1, X_2 be the random variables corresponding to the above roots, then

$$X_1=\frac{a+\sqrt{a^2-4Y}}{2},\ X_2=\frac{a-\sqrt{a^2-4Y}}{2},$$

where Y is the random variable corresponding to b.

Here $a^2-4b \geqslant 0$ and $b>0$. So we have $0<b\leqslant\frac{a^2}{4}$.

Hence, the random variable Y has uniform distribution in $\left(0,\frac{a^2}{4}\right)$. So, if $f_Y(b)$ be the probability density function of Y, then

$$f_Y(b)=\frac{1}{\frac{a^2}{4}} \quad \text{if } 0<b<\frac{a^2}{4}$$

$$=0\ , \quad \text{elsewhere.}$$

$$\therefore\ E(X_1)=\int_0^{\frac{a^2}{4}} \frac{a+\sqrt{a^2-4b}}{2}\cdot\frac{4}{a^2}\,db$$

$$=\frac{2}{a^2}\left[ab+\tfrac{2}{3}(a^2-4b)^{\frac{3}{2}}\left(-\tfrac{1}{4}\right)\right]_0^{\frac{a^2}{4}}$$

$$=\frac{2}{a^2}\left(\frac{a^3}{4}+\tfrac{1}{6}a^3\right)=\frac{5a}{6}.$$

Similarly, we can show that $E(X_2)=\frac{a}{6}$.

So the required mean values are $\frac{5a}{6}, \frac{a}{6}$.

Ex. 12. *The radius X of a circle has uniform distribution in (1, 2). Find the mean and variance of the area of the circle.*

If Y be the random variable corresponding to the area of the circle, then

$$Y=\pi X^2.$$

The probability density function of X is given by

$$f(x)=1,\ \text{if } 1<x<2$$

$$=0,\ \text{elsewhere.}$$

Now $E(Y)=E(\pi X^2)=\pi E(X^2)$

$$=\pi\int_1^2 x^2\,dx=\frac{7\pi}{3}.$$

$$\text{Var }(Y)=E(Y^2)-\{E(Y)\}^2$$
$$=E(\pi^2X^4)-\left(\frac{7\pi}{3}\right)^2$$
$$=\pi^2\int_1^2 x^4\,dx-\frac{49}{9}\pi^2$$
$$=\frac{31\,\pi^2}{5}-\frac{49\,\pi^2}{9}$$
$$=\frac{34\,\pi^2}{45}.$$

Ex. 13. *Two players, Sunil and Kapil, agree to play a game under the condition that Sunil will get from Kapil R_1 rupees if he wins and will pay Kapil R_1 rupees if he loses. The probability of Sunil's winning is p. Denote Sunil's gain by Z. Show that the variance of Z (i) increases with R_1,*

(ii) decreases as p increases for $p > \frac{1}{2}$. [C. H. (Econ.) '90]

Here the spectrum of Z is $\{R_1, -R_1\}$.

Also it is given that $P\,(Z=R_1)=p$,

$$P\,(Z=-R_1)=1-p.$$

Then $E(Z)=R_1 \cdot p+(-R_1)\,(1-p)=(2p-1)\,R_1$.

$$\text{Var }(Z)=E(Z^2)-\{E(Z)\}^2$$
$$=R_1{}^2p+(-R_1)^2\,(1-p)-(2p-1)^2\,R_1{}^2$$
$$=4p\,(1-p)\,R_1{}^2,$$

which shows that Var (Z) increases with R_1, since $p\,(1-p)>0$. So (i) is proved.

Again, we see that

Var $(Z)=R_1{}^2\,\{1-(2p-1)^2\}=\psi\,(p)$, say.

$\psi'(p)=-4(2p-1)\,R_1{}^2$, R_1 being kept fixed.

$\therefore\ \psi'(p)<0$ if $2p-1>0$, *i.e.*, $p>\frac{1}{2}$.

Hence, $\psi(p)$, *i.e.*, Var (Z) decreases as p increases when $p>\frac{1}{2}$.

Ex. 14. *Find the median and mode (if any) of the continuous distribution with probability density function given by*

$$f(x)=k(x-9)(10-x) \text{ if } 9 \leqslant x \leqslant 10$$
$$=0\text{, elsewhere.}$$

$f(x)$ being a probability density function, we have

$$\int_{9}^{10} k(x-9)\,(10-x)\,dx=1$$

$$\text{or,}\quad k\int_{0}^{1} t(1-t)\,dt=1 \qquad \text{where}\quad t=x-9$$

$$\text{or,}\quad k\left(\tfrac{1}{2}-\tfrac{1}{3}\right)=1. \qquad \therefore\quad k=6.$$

If $F(x)$ be the distribution function of the corresponding random variable, then

$$\begin{aligned} F(x)&=0 \quad \text{if} \quad -\infty < x < 9\\ &=\int_{9}^{x} 6(t-9)\,(10-t)\,dt \quad \text{if} \quad 9 \leqslant x \leqslant 10\\ &=1 \quad \text{if} \quad x > 10. \end{aligned}$$

$$\text{Now,}\quad \int_{9}^{x} 6(t-9)\,(10-t)\,dt=6\int_{0}^{x-9} u(1-u)\,du \qquad \text{when } t-9=u$$

$$=6\left\{\frac{(x-9)^2}{2}-\frac{(x-9)^3}{3}\right\}$$

$$=(x-9)^2\,(21-2x).$$

$$\text{Now,}\quad F(x)\neq\tfrac{1}{2} \quad \text{in} \quad -\infty < x < 9$$

$$\neq\tfrac{1}{2} \quad \text{in } x > 10.$$

If $9 \leqslant x \leqslant 10$, then $F(x)=\frac{1}{2}$ gives

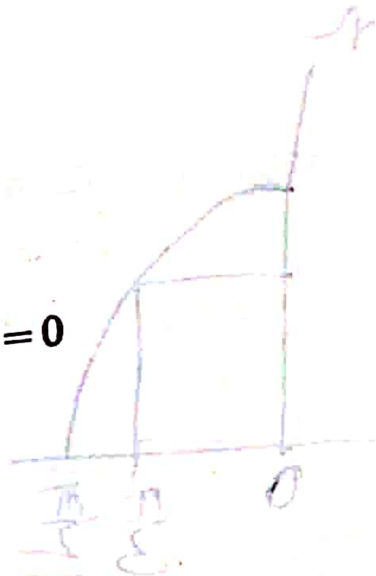
$$(x-9)^2\,(21-2x)=\tfrac{1}{2}$$

$$\text{or,}\quad 4x^3-114x^2+1080x-3401=0$$

$$\text{or,}\quad 2x^2(2x-19)-38x(2x-19)+179(2x-19)=0$$

$$\text{or,}\quad (2x-19)\,(2x^2-38x+179)=0.$$

$$\therefore\quad x=\frac{19}{2},\ \frac{19\pm\sqrt{3}}{2}.$$



But $\dfrac{19+\sqrt{3}}{2} > 10$ and $\dfrac{19-\sqrt{3}}{2} < 9$.

$\therefore$ $x=\frac{19}{2}$ is the only solution of $F(x)=\frac{1}{2}$ in $9 \leqslant x \leqslant 10$. So the required median is $\frac{19}{2}$.

Now $f'(x)=6(19-2x)$ in $9 < x < 10$.

Then $f'(x)=0$ in $9 < x < 10$ gives $x=\frac{19}{2}$.

Again, $f''(x) = -12$ in $9 < x < 10$

$\therefore \quad f''(\frac{19}{2}) = -12 < 0.$

Hence, $f(x)$ has a maximum at $x = \frac{19}{2}$.

So the given distribution has a unique mode and it is $\frac{19}{2}$.

Ex. 15. *Find the median and mode (if any) of the continuous distribution with probability density function given by*

$$f(x) = \sin x \quad \text{if } 0 \leqslant x \leqslant \frac{\pi}{2}$$
$$= 0 \quad , \quad \textit{elsewhere.}$$

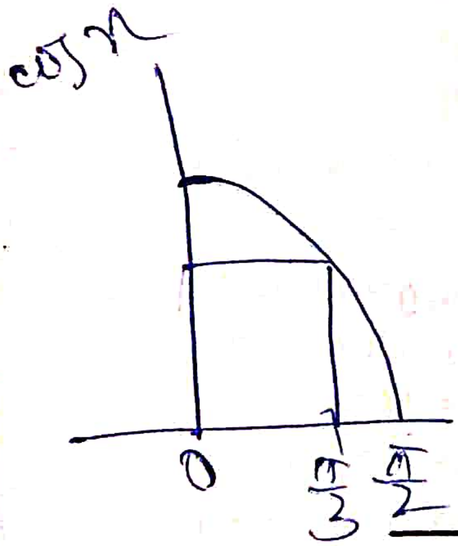
The corresponding distribution function $F(x)$ is given by

$$F(x) = 0 \quad \text{if } -\infty < x < 0$$
$$= \int_0^x \sin t \, dt \text{ if } 0 \leqslant x \leqslant \frac{\pi}{2}$$
$$= 1 \quad \text{if } x > \frac{\pi}{2}.$$

Now, $F(x) = \frac{1}{2}$ gives $\int_0^x \sin t \, dt = \frac{1}{2}$ where $0 \leqslant x \leqslant \frac{\pi}{2}$.

or, $-\cos x + 1 = \frac{1}{2}$, *i.e.*, $\cos x = \frac{1}{2}$,

which has the only solution $x = \frac{\pi}{3}$ satisfying $0 \leqslant x \leqslant \frac{\pi}{2}$.



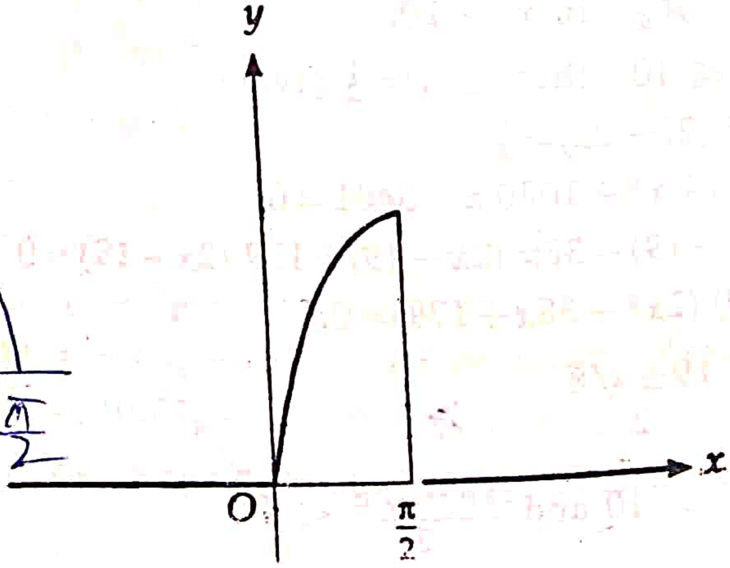


Fig. 7.8.2

Hence, the required median is $\frac{\pi}{3}$.

Fig. 7.8.2 represents the graph of $y=f(x)$.

Now in $0 < x < \frac{\pi}{2}$, $f'(x)=0$ gives $\cos x=0$, which has no solution in $0 < x < \frac{\pi}{2}$. So $f(x)$ has no local maximum in

$$0 < x < \frac{\pi}{2}.$$

Now, $f\left(\frac{\pi}{2}\right)=1$, $f(x)=0 < 1$ for all $x > \frac{\pi}{2}$,

$$f(x) < 1 \text{ in } 0 < x < \frac{\pi}{2}.$$

So, we can find a positive number δ such that

$$f(x) < f\left(\frac{\pi}{2}\right) \text{ for } \frac{\pi}{2}-\delta < x < \frac{\pi}{2}+\delta,\ x \neq \frac{\pi}{2}.$$

So, $f(x)$ has a local maximum at $x=\frac{\pi}{2}$.

Hence, the given distribution has a unique mode at $x=\frac{\pi}{2}$.

Ex. 16. *Find the median and mode of a normal (m, σ) distribution.*

The probability density function of a normal (m, σ) variate is given by

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

The distribution function of the corresponding random variable X is given by

$$F(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\,dt.$$

Now the median μ is the least value of x for which $F(x)=\frac{1}{2}$.

Now $F(x)=\frac{1}{2}$ gives

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\,dt=\tfrac{1}{2}.$$

$$\text{or,} \quad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}}\,du=\tfrac{1}{2}, \quad \text{where } u=\frac{t-m}{\sigma}. \qquad (7.8.2)$$

Case I. Let $x > m$. From (7.8.2) we get

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{u^2}{2}}\,du+\frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}}\,du=\tfrac{1}{2}$$

or, $$\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_{0}^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}}\,du=\tfrac{1}{2}$$

or, $$\int_{0}^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}}\,du=0. \tag{7.8.3}$$

But $e^{-\frac{u^2}{2}}$ is continuous in $\left[0, \frac{x-m}{\sigma}\right]$ and $e^{-\frac{u^2}{2}} > 0$ for all $u \in \left[0, \frac{x-m}{\sigma}\right]$ and hence $\int_0^{\frac{x-m}{\sigma}} e^{-\frac{u^2}{2}} > 0$ and this is contrary to (7.8.3). So, there is no value of x satisfying $F(x)=\frac{1}{2}$ for $x > m$.

Case II. Let $x < m$.

Here (7.8.2) gives

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{u^2}{2}}\,du-\frac{1}{\sqrt{2\pi}}\int_{\frac{x-m}{\sigma}}^{0} e^{-\frac{u^2}{2}}\,du=\tfrac{1}{2}$$

or, $$\int_{\frac{x-m}{\sigma}}^{0} e^{-\frac{u^2}{2}}\,du=0,$$

but $\int_{\frac{x-m}{\sigma}}^{0} e^{-\frac{u^2}{2}} > 0$, since $e^{-\frac{u^2}{2}}$ is continuous and positive in $\left[\frac{x-m}{\sigma}, 0\right]$. So, $F(x)=\frac{1}{2}$ has no soultion for $x < m$.

Lastly, we find that for $x=m$,

$$F(m)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-\frac{u^2}{2}}\,du=\tfrac{1}{2}.$$

So, m is the only solution of the equation $F(x)=\frac{1}{2}$.
Hence, m is the median of the normal (m, σ) distribution.
Again, $f'(x)$ exists for all $x \in (-\infty, \infty)$.

Now, $f'(x)=0$ gives, $-e^{-\frac{(x-m)^2}{2\sigma^2}} \cdot \frac{x-m}{\sigma^2}=0$, *i.e.*, $x=m$.

$$\text{Again, } f''(x)=-\frac{1}{\sigma^2}\, e^{-\frac{(x-m)^2}{2\sigma^2}}+\frac{(x-m)^2}{\sigma^4}\, e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

$$\therefore \quad f''(m)=-\frac{1}{\sigma^2}<0.$$

So, $f(x)$ has a unique local maximum at $x=m$.
So, m is the unique mode of the normal (m, σ) distribution.

Ex. 17. *The probability density function of a continuous random variable X is given by*

$$f(x)=\frac{1}{(1+x)^2} \quad \textit{if } 0 \leqslant x < \infty$$
$$=0, \quad \textit{elsewhere}.$$

Investigate whether the mean and the variance of the distribution exist. Also find the median and quartiles of the distribution.

The mean of the distribution exists if $\int_0^\infty \frac{x}{(1+x)^2}\, dx$ is absolutely convergent.

$$\text{Now, } \underset{B\to\infty}{Lt} \int_0^B \left| \frac{x}{(1+x)^2} \right| dx,\ B>0$$

$$=\underset{B\to\infty}{Lt} \int_0^B \frac{x}{(1+x)^2}\, dx=\underset{B\to\infty}{Lt} \int_1^{1+B} \frac{u-1}{u^2}\, du, \qquad \text{where } 1+x=u$$

$$=\underset{B\to\infty}{Lt} \left\{ \log\,(B+1)+\frac{1}{B+1}-1 \right\}$$

$$=\infty.$$

Hence, $\int_0^\infty \frac{x}{(1+x)^2}\, dx$ is not absolutely convergent. So the mean does not exist and consequently the variance also does not exist.

Les $F(x)$ be the distribution function of the corresponding random variable X.

Then $F(x)=\int_{-\infty}^{x} f(t)\, dt=0$ if $x<0$

$$=\int_{0}^{x} \frac{dt}{(1+t)^2}\, dt \text{ if } x \geqslant 0$$

i.e., $F(x)=0$ if $x<0$

$$=\frac{x}{1+x} \text{ if } x \geqslant 0.$$

Now, $F(x)=\frac{1}{2}$ gives $\frac{x}{1+x}=\frac{1}{2}$ *i.e.*, $x=1$.

Hence the median is 1.

Again $F(x)=\frac{3}{4}$, $F(x)=\frac{1}{4}$ give respectively, $x=3$ and $x=\frac{1}{3}$ which give the quartiles of the distribution.

Ex. 18. *The probability density function of a continuous random variable X is given by*

$$f(x)=ke^{-b(x-a)}, \text{ if } a \leqslant x < \infty$$
$$=0 \quad , \text{ elsewhere,}$$

where k, a, b (> 0) are constants. Show that $k=b=\sigma^{-1}$ *and* $a=m-\sigma$, *where m, σ are the mean and the standard deviation of the distribution. Also find* γ_1 *and* γ_2.

Here $k\int_{a}^{\infty} e^{-b(x-a)}\, dx=1$ or, $\frac{k\, e^{ab}.\, e^{-ab}}{b}=1$, *i.e.*, $k=b$.

Now, $m=E(X)=b\int_{a}^{\infty} x\, e^{-b(x-a)}\, dx$

$$=b\int_{0}^{\infty} (a+z)\, e^{-bz}\, dz, \quad \text{where } x-a=z$$

$$=ab\int_{0}^{\infty} e^{-bz}\, dz+b\int_{0}^{\infty} ze^{-bz}\, dz$$

$$=\frac{ab}{b}+\frac{1}{b}\int_{0}^{\infty} ue^{-u}\, du, \quad \text{where } bz=u$$

$$=a+\frac{1}{b}\,\Gamma(2)=a+\frac{1}{b}.$$

Again, $E(X^2)=b\int_a^{\infty} x^2\, e^{-b(x-a)}\, dx$

$$=b\int_0^{\infty} (a+z)^2\, e^{-bz}\, dz$$

$$=ba^2\int_0^{\infty} e^{-bz}\, dz+2ab\int_0^{\infty} z\, e^{-bz} dz+b\int_0^{\infty} z^2\, e^{-bz} dz$$

$$=\frac{ba^2}{b}+\frac{2a}{b}+\frac{1}{b^2}\int_0^{\infty} u^2 e^{-u}\, du$$

$$=a^2+\frac{2a}{b}+\frac{1}{b^2}\,\Gamma(3)$$

$$=a^2+\frac{2a}{b}+\frac{2}{b^2}\cdot$$

$\therefore\quad \sigma^2=\text{Var}\,(X)=E(X^2)-\{E(X)\}^2$

$$=a^2+\frac{2a}{b}+\frac{2}{b^2}-\left(a+\frac{1}{b}\right)^2=\frac{1}{b^2}\cdot$$

$\therefore\quad \sigma=\frac{1}{b},\ k=b=\sigma^{-1}$ and $m=a+\sigma.\quad i.e.,\ a=m-\sigma.$

Now, $\mu_3=E\{(X-m)^3\}$

$$=E(X^3)-3m\,E(X^2)+3m^2\,E(X)-m^3$$

$$=E(X^3)-3m\left(a^2+\frac{2a}{b}+\frac{2}{b^2}\right)+3m^3-m^3$$

$$=E(X^3)-3m\left\{\left(a+\frac{1}{b}\right)^2+\frac{1}{b^2}\right\}+2m^3$$

$$=E(X^3)-3m\,(m^2+\sigma^2)+2m^3$$

$$=E(X^3)-m^3-3m\sigma^2.$$

Now, $E(X^3)=b\int_a^{\infty} e^{-b(x-a)}\, x^3\, dx$

$$=b\int_0^{\infty} e^{-bz}\,(z+a)^3\, dz$$

$$=b\int_0^{\infty} z^3 e^{-bz}\, dz+3ab\int_0^{\infty} z^2 e^{-bz}\, dz$$

$$+3a^2b\int_0^{\infty} ze^{-bz}\, dz+ba^3\int_0^{\infty} e^{-bz}\, dz$$

or, $$E(X^3)=\frac{1}{b^3}\,\Gamma(4)+\frac{3ab}{b^3}\,\Gamma(3)+\frac{3a^2b}{b^2}\,\Gamma(2)+a^3$$

$$=\frac{6}{b^3}+\frac{6a}{b^2}+\frac{3a^2}{b}+a^3$$

$$=a^3+\frac{3}{b}\left(a+\frac{1}{b}\right)^2+\frac{3}{b^3}$$

$$=a^3+3\sigma m^2+3\sigma^3$$

$$\therefore\quad \mu_3=a^3+3\sigma m^2+3\sigma^3-m^3-3m\sigma^2$$

$$=(m-\sigma)^3+3\sigma m^2+3\sigma^3-m^3-3m\sigma^2$$

$$=2\sigma^3.$$

$$\therefore\quad \gamma_1=\frac{\mu_3}{\sigma^3}=\frac{2\sigma^3}{\sigma^3}=2.$$

Again, $\mu_4=E\{(X-m)^4\}$

$$=E(X^4)-4mE(X^3)+6m^2E(X^2)-4m^3E(X)+m^4$$

$$=E(X^4)-4m(a^3+3\sigma m^2+3\sigma^3)+6m^2(m^2+\sigma^2)-4m^4+m^4$$

$$=E(X^4)-12m^3\sigma+6m^2\sigma^2-12m\sigma^3-4a^3m+3m^4.$$

Now, $$E(X^4)=b\int_a^\infty e^{-b(x-a)}\,x^4\,dx$$

$$=b\int_0^\infty e^{-bz}\,(z+a)^4\,dz$$

$$=b\int_0^\infty z^4\,e^{-bz}\,dz+4ab\int_0^\infty z^3\,e^{-bz}\,dz$$

$$+6a^2b\int_0^\infty z^2\,e^{-bz}dz+4a^3b\int_0^\infty ze^{-bz}+a^4b\int_0^\infty e^{-bz}dz$$

$$=\frac{1}{b^4}\,\Gamma(5)+\frac{4ab}{b^4}\,\Gamma(4)+\frac{6a^2b}{b^3}\,\Gamma(3)+\frac{4a^3b}{b^2}\,\Gamma(2)+a^4$$

$$=\frac{24}{b^4}+\frac{24a}{b^3}+\frac{12a^2}{b^2}+\frac{4a^3}{b}+a^4$$

$$=\frac{4}{b}\left(a+\frac{1}{b}\right)^3+\frac{12}{b^3}\left(a+\frac{1}{b}\right)+\frac{8}{b^4}+a^4$$

$$=4m^3\sigma+12m\sigma^3+8\sigma^4+a^4.$$

$$\therefore \quad \mu_4 = a^4 + 4m^3\sigma + 12m\sigma^3 + 8\sigma^4 - 12m^3\sigma + 6m^2\sigma^2$$
$$- 12m\sigma^3 - 4a^3m + 3m^4$$
$$= (m-\sigma)^4 - 8m^3\sigma + 6m^2\sigma^2 + 8\sigma^4 - 4(m-\sigma)^3 m + 3m^4$$
$$= 9\sigma^4.$$

$$\therefore \quad \gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = 6.$$

Ex. 19. *Find the median and the quartiles of Cauchy distribution with parameters λ, μ.*

The distribution function $F(x)$ of the random variable X having Cauchy distribution with parameters (λ, μ) is given by

$$F(x) = \frac{\lambda}{\pi}\int_{-\infty}^{x} \frac{dt}{(t-\mu)^2 + \lambda^2} = \frac{\lambda}{\pi}\left[\frac{\tan^{-1}\left(\frac{t-\mu}{\lambda}\right)}{\lambda}\right]_{-\infty}^{x}$$

$$= \frac{1}{\pi}\tan^{-1}\frac{x-\mu}{\lambda} + \tfrac{1}{2}. \quad = \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\lambda}\right) + \frac{1}{\pi}$$

If x_1 be the median of X, then $F(x_1) = \frac{1}{2}$.

$$\therefore \quad \frac{1}{\pi}\tan^{-1}\frac{x_1-\mu}{\lambda} + \tfrac{1}{2} = \tfrac{1}{2},$$

i.e., $x_1 = \mu$,

which is the required median.

The upper quartile $\zeta_{\frac{3}{4}}$ is given by $F(\zeta_{\frac{3}{4}}) = \frac{3}{4}$.

$$\therefore \quad \frac{1}{\pi}\tan^{-1}\frac{\zeta_{\frac{3}{4}}-\mu}{\lambda} + \tfrac{1}{2} = \tfrac{3}{4}.$$

$$\therefore \qquad \zeta_{\frac{3}{4}} = \mu + \lambda.$$

Similarly, we find that $\zeta_{\frac{1}{4}} = \mu - \lambda$.

Ex. 20. *Find the median of binomial* $(5, \frac{1}{2})$ *distribution.*

The distribution function $F(x)$ of the random variable X having binomial $(5, \frac{1}{2})$ distribution is given by

$$F(x) = 0 \text{ if } -\infty < x < 0$$
$$= (\tfrac{1}{2})^5 \text{ if } 0 \leqslant x < 1$$
$$= (\tfrac{1}{2})^5 + {}^5C_1(\tfrac{1}{2})^5 \text{ if } 1 \leqslant x < 2$$

$$=(\tfrac{1}{2})^5+{}^5C_1(\tfrac{1}{2})^5+{}^5C_2(\tfrac{1}{2})^5 \text{ if } 2 \leqslant x < 3$$
$$=(\tfrac{1}{2})^5+{}^5C_1(\tfrac{1}{2})^5+{}^5C_2(\tfrac{1}{2})^5+{}^5C_3(\tfrac{1}{2})^5 \text{ if } 3 \leqslant x < 4$$
$$=(\tfrac{1}{2})^5+{}^5C_1(\tfrac{1}{2})^5+{}^5C_2(\tfrac{1}{2})^5+{}^5C_3(\tfrac{1}{2})^5+{}^5C_4(\tfrac{1}{2})^5 \text{ if } 4 \leqslant x < 5$$
$$=1 \text{ if } x \geqslant 5.$$

We see that $F(x)=\frac{1}{2}$ and $F(x-0)=\frac{1}{2}$ for all $x \in (2, 3)$

and $F(2)=\frac{1}{2},\ F(2-0)=\frac{6}{32} < \frac{1}{2}$,

$F(3)=\frac{13}{16} > \frac{1}{2},\ F(3-0)=\frac{1}{2}$.

So $F(x-0) \leqslant \frac{1}{2}$ for $x \in [2, 3]$

and $F(x) \geqslant \frac{1}{2}\ \forall\ x \in [2, 3]$.

So any value of x belonging to $[2, 3]$ satisfies simultaneously $F(x) \geqslant \frac{1}{2}$, $F(x-0) \leqslant \frac{1}{2}$. So here the median is $\frac{1}{2}(2+3)=2{\cdot}5$.

Ex. 21. *Find the mode or modes of binomial (n, p) distribution.*

Let X be a binomial (n, p) variate. The probability mass function of X is given by

$$f(x)={}^nC_x\, p^x\,(1-p)^{n-x} \text{ for } x=0, 1, 2, \ldots n.$$

Now, $f(x) \geqslant f(x+1)$ if

$${}^nC_x p^x (1-p)^{n-x} \geqslant {}^nC_{x+1} p^{x+1} (1-p)^{n-x-1}$$

or, $$\frac{1-p}{n-x} \geqslant \frac{p}{x+1} \quad (x \neq n)$$

or, $x \geqslant (n+1)\,p-1$.

Similarly, we find that $f(x) \geqslant f(x-1)$ if $x \leqslant (n+1)\,p$.

Case I. Let $(n+1)\,p$ be an integer.

In this case, $f(M-2) < f(M-1) = f(M) > f(M+1)$, where $M=(n+1)p$. So here $(n+1)\,p-1$, $(n+1)\,p$ are the modes of binomial (n, p) distribution.

Case II. Let $M=(n+1)p$ be not an integer. Then we have $M-1 < [M] < M$, where $[M]$ is the greatest integer not greater than M and then

$$f([M]) > f([M]+1)$$

and $$f([M]) > f([M]-1).$$

Hence $[M]=[(n+1)p]$ is the unique mode of the distribution when $(n+1)p$ is not an integer.

Ex. 22. *Find the mode or modes of Poisson distribution with parameter μ.*

The probability mass function of Poisson μ variate is given by

$$f(x)=\frac{e^{-\mu}\mu^x}{x!} \quad \text{for } x=0, 1, 2, \ldots$$

Now, $f(x) \geqslant f(x+1)$ if $\dfrac{e^{-\mu}\mu^x}{x!} \geqslant \dfrac{e^{-\mu}\mu^{x+1}}{(x+1)!}$,

i.e., if $x \geqslant \mu-1$.

Again, $f(x) \geqslant f(x-1)$ if $\dfrac{e^{-\mu}\mu^x}{x!} \geqslant \dfrac{e^{-\mu}\mu^{x-1}}{(x-1)!}$,

i.e., if $x \leqslant \mu$.

Case I. Let μ be an integer. In this case,

$$f(\mu-1)=f(\mu) > f(\mu+1)$$

and $f(\mu)=f(\mu-1) > f(\mu-2)$,

where $\mu-1$, $\mu-2$, μ are non-negative integers. So in this case we have

$$f(\mu-2) < f(\mu-1)=f(\mu) > f(\mu+1).$$

Hence, $\mu-1$, μ are modes of Poisson μ distribution when μ is an integer.

Case II. Let μ be not an integer. In this case, we have $\mu-1 < [\mu] < \mu$, where $[\mu]$ is the greatest integer not greater than μ and then

$$f([\mu]) > f([\mu]+1) \text{ and}$$
$$f([\mu]) > f([\mu]-1).$$

Hence, $[\mu]$ is the unique mode of Possion μ distribution when μ is not an integer.

Ex. 23. *If $\log X$ has normal distribution with $m=1$, $\sigma=2$, then find $E(X)$, $Var(X)$.*

Let $Y=\log X$. Then $X=e^Y$ where Y is a normal (1, 2) variate. If $M_Y(t)$ be the moment generating functi on of Y, then

$$M_Y(t)=e^{t+2t^2}=E(e^{tY}).$$

[The moment generating function of a normal (m, σ) variate is $e^{mt+\frac{1}{2}\sigma^2 t^2}$.]

Now, $E(X)=E(e^Y)=M_Y(1)=e^3$. $[\because M_Y(t)= E(e^{tY})]$

Again, $\text{Var}(X)=E(X^2)-[E(X)]^2$

$$=E(e^{2Y})-e^6$$

$$=e^{10}-e^6.$$

**** **Ex. 24.** *Show that the mean deviation about the mean of a normal (m, σ) distribution is $\sqrt{\frac{2}{\pi}}\,\sigma$.*

The required mean deviation is $E(|X-m|)$, where X is a normal (m, σ) variate.

$$\text{Now, } E(|X-m|)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}|x-m|\,e^{-\frac{(x-m)^2}{2\sigma^2}}dx$$

i.e. $|x-m|$ is even function

$$=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}|y|\,e^{-\frac{y^2}{2\sigma^2}}dy,\ (x-m=y)$$

$$=\frac{2}{\sqrt{2\pi}\sigma}\int_{0}^{\infty}|y|\,e^{-\frac{y^2}{2\sigma^2}}dy,$$

since $\int_0^\infty |y|\,e^{-\frac{y^2}{2\sigma^2}}dy$ is convergent and $|y|\,e^{-\frac{y^2}{2\sigma^2}}$ is an even function.

$$\text{Then } E(|X-m|)=\sqrt{\frac{2}{\pi}}\int_0^\infty \frac{y}{\sigma}\,e^{-\frac{y^2}{2\sigma^2}}dy$$

$$=\sqrt{\frac{2}{\pi}}\,\sigma\int_0^\infty z e^{-\frac{z^2}{2}}dz,\ \text{ where } y=\sigma z.$$

$$=\sqrt{\frac{2}{\pi}}\,\sigma\int_0^\infty e^{-t}dt,\ \text{ where } \frac{z^2}{2}=t.$$

$$=\sqrt{\frac{2}{\pi}}\,\sigma\,\Gamma(1)=\sqrt{\frac{2}{\pi}}\,\sigma.$$

Ex. 25. *Find the median and mode of a binomial* $(4, \frac{1}{4})$ *variate.*

The distribution function of the binomial $(4, \frac{1}{4})$ variate is given by

$$F(x) = 0 \text{ if } -\infty < x < 0$$
$$= (\tfrac{3}{4})^4 \text{ if } 0 \leqslant x < 1$$
$$= (\tfrac{3}{4})^4 + {}^4C_1 \tfrac{1}{4}(\tfrac{3}{4})^3 \text{ if } 1 \leqslant x < 2$$
$$= (\tfrac{3}{4})^4 + {}^4C_1 \tfrac{1}{4}(\tfrac{3}{4})^3 + {}^4C_2(\tfrac{1}{4})^2(\tfrac{3}{4})^2 \text{ if } 2 \leqslant x < 3$$
$$= (\tfrac{3}{4})^4 + {}^4C_1 \tfrac{1}{4}(\tfrac{3}{4})^3 + {}^4C_2(\tfrac{1}{4})^2(\tfrac{3}{4})^2 + {}^4C_3(\tfrac{1}{4})^3(\tfrac{3}{4}) \text{ if } 3 \leqslant x < 4$$
$$= 1 \quad \text{if } x \geqslant 4.$$

It follows that $F(1) > \frac{1}{2}$ and $F(1-0) < \frac{1}{2}$ and further we note that no other value of x satisfies simultaneously the inequalities $F(x) \geqslant \frac{1}{2}$ and $F(x-0) \leqslant \frac{1}{2}$. Hence the required median is 1.

Here $n=4$, $p=\frac{1}{4}$ and $(n+1)\,p=\frac{5}{4}$, which is not an integer. So, by Ex. 21, the mode is the unique integer belonging to $(\frac{5}{4}-1, \frac{5}{4})$, *i.e.*, $(\frac{1}{4}, \frac{5}{4})$ and this integer is 1. So the mode is also 1.

Ex. 26. *Find the median and mode of the Poisson distribution with mean* 2.

Let X be a Poisson variate with mean $\mu=2$. The probability mass function of X is given by $f(x) = \dfrac{e^{-2}2^x}{x\,!}$ for $x = 0, 1, 2, \ldots\ldots$

The distribution function $F(x)$ of X is given by

$$F(x) = 0 \qquad \text{if } -\infty < x < 0$$
$$= e^{-2} \qquad \text{if } 0 \leqslant x < 1$$
$$= e^{-2} + \frac{e^{-2}\,.\,2}{1\,!} \qquad \text{if } 1 \leqslant x < 2$$
$$= \ldots\ldots$$

Now, $e^{-2} + e^{-2}\,.\,\dfrac{2}{1\,!} = \dfrac{3}{e^2}$, and

$$\frac{3}{e^2} - \frac{1}{2} = \frac{6-e^2}{2e^2},$$

$$e > 1 + \frac{1}{1\,!} + \frac{1}{2\,!} = \tfrac{5}{2}. \quad \therefore \quad e^2 > \tfrac{25}{4}.$$

$\therefore\ 6-e^2 < 6-\frac{25}{4} < 0.$

$\therefore\ \frac{3}{e^2} < \frac{1}{2}.$

So, $F(x)=\frac{3}{e^2} < \frac{1}{2}$ for $x \in [1, 2)$.

Again, $e^{-2}+2e^{-2}+\frac{e^{-2}.\,2^2}{2\,!}=\frac{5}{e^2}$.

Now, $e < 3$. $\therefore\ \frac{5}{e^2} > \frac{5}{9} > \frac{1}{2}$.

Then, $F(x)=\frac{5}{e^2} > \frac{1}{2}$ for $x \in [2, 3)$.

Hence, $F(2-0)=\frac{3}{e^2} < \frac{1}{2}$ and $F(2)=\frac{5}{e^2} > \frac{1}{2}$.

So, $x=2$ satisfies simultaneously $F(x) \geqslant \frac{1}{2}$ and $F(x-0) \leqslant \frac{1}{2}$ and we note that no other value of x satisfies these inequalities. So,2 is the unique median of the given distribution.

Here $\mu=2$ is an integer. So by *Ex.* 22, $\mu-1, \mu$, *i.e.*, 1, 2 are the two modes of the given distribution.

$|x-m|$

Ex. 27. *Prove that for any distribution, the first absolute moment about the mean cannot exceed the standard deviation.*

Let X be a random variable where $E(X)=m$ and Var $(X)=\sigma^2$. We have Var $(X)=E(X^2)-\{E(X)\}^2$.

Now Var $(X)=E(X-m)^2 \geqslant 0$. $\therefore\ E(X^2) \geqslant \{E(X)\}^2$ for any random variable X for which $E(X)$, $E(X^2)$ exist.

Let $Y=|X-m|$. Then by the above inequality,

$$E(Y^2) \geqslant \{E(Y)\}^2$$

or, $$E(|X-m|^2) \geqslant \{E(|X-m|)\}^2$$

or, $$E(X-m)^2 \geqslant \{E(|X-m|)\}^2$$

or, $$\text{Var}(X) \geqslant \{E(|X-m|)\}^2.$$

Hence, $\sigma \geqslant E(|X-m|)$, *i.e.*, $E(|X-m|) \leqslant \sigma$.

Hence,it is proved that the first absolute moment of X about the mean cannot exceed the standard deviation of X.

Ex. 28. *The probability density of a continuous random variable X is given by*

$$f(x)=\frac{2(b+x)}{b(a+b)},\quad -b \leqslant x < 0$$

$$=\frac{2(a-x)}{a(a+b)},\quad 0 \leqslant x \leqslant a.$$

Find the mean, variance and median of the distribution of X, if $a > b > 0$.

The mean m is given by

$$m=E(X)=\int_{-b}^{0} \frac{2(b+x)x}{b(a+b)}\,dx+\int_{0}^{a} \frac{2(a-x)x}{a(a+b)}\,dx$$

$$=\frac{2}{b(a+b)}\left(-\frac{b^3}{2}+\frac{b^3}{3}\right)+\frac{2}{a(a+b)}\left(\frac{a^3}{2}-\frac{a^3}{3}\right)$$

$$=-\frac{b^2}{3(a+b)}+\frac{a^2}{3(a+b)}$$

$$=\tfrac{1}{3}(a-b).$$

Now, $$E(X^2)=\int_{-b}^{0} \frac{2(b+x)}{b(a+b)}x^2\,dx+\int_{0}^{a} \frac{2(a-x)}{a(a+b)}\,x^2\,dx$$

$$=\frac{2}{a+b}\left(\frac{b^3}{3}-\frac{b^3}{4}\right)+\frac{2}{a+b}\left(\frac{a^3}{3}-\frac{a^3}{4}\right)$$

$$=\frac{b^3}{6(a+b)}+\frac{a^3}{6(a+b)}=\tfrac{1}{6}(a^2-ab+b^2).$$

$\therefore$ $$\text{Var}(X)=E(X^2)-m^2$$

$$=\tfrac{1}{6}(a^2-ab+b^2)-\tfrac{1}{9}(a-b)^2$$

$$=\tfrac{1}{18}(a^2+ab+b^2).$$

The distribution function $F(x)$ of X is given by

$$F(x)=0 \text{ if } -\infty < x < -b$$

$$=\int_{-b}^{x} \frac{2(b+t)}{b(a+b)}\,dt \text{ if } -b \leqslant x < 0$$

$$=\int_{-b}^{0} \frac{2(b+t)}{b(a+b)}\,dt+\int_{0}^{x} \frac{2(a-t)}{a(a+b)}\,dt \text{ if } 0 \leqslant x \leqslant a$$

$$=1 \text{ if } x > a.$$

Now, $\int_{-b}^{x} \frac{2(b+t)}{b(a+b)}\,dt = \frac{2}{b(a+b)}\left(bx+\frac{x^2}{2}+\frac{b^2}{2}\right)$,

and $\int_{0}^{x} \frac{2(a-t)}{a(a+b)}\,dt = \frac{2}{a(a+b)}\left(ax-\frac{x^2}{2}\right)$.

$$\therefore\ F(x) = 0 \quad \text{if } -\infty < x < -b$$
$$= \frac{2}{b(a+b)}\left(bx+\frac{x^2}{2}+\frac{b^2}{2}\right) \quad \text{if } -b \leqslant x < 0$$
$$= \frac{b}{a+b}+\frac{2}{a(a+b)}\left(ax-\frac{x^2}{2}\right) \quad \text{if } 0 \leqslant x \leqslant a$$
$$= 1 \quad \text{if } x > a.$$

Now, $F(x) \neq \frac{1}{2}$ if $-\infty < x < -b$.

If $-b \leqslant x < 0$, then $F(x) = \frac{1}{2}$ gives

$$\frac{(b+x)^2}{b(a+b)} = \frac{1}{2}.$$

$$\therefore\ x = -b \pm \sqrt{\frac{b(a+b)}{2}}.$$

But $-b-\sqrt{\frac{b(a+b)}{2}} < -b$ and if $a > b$, then

$$-b+\sqrt{\frac{b(a+b)}{2}} > -b+b = 0.$$

$$\therefore\ \text{if } a > b,\ -b \pm \sqrt{\frac{b(a+b)}{2}} \notin [-b, 0).$$

Now, if $0 \leqslant x \leqslant a$, then $F(x) = \frac{1}{2}$ gives

$$\frac{b}{a+b}+\frac{2}{a(a+b)}\left(ax-\frac{x^2}{2}\right) = \frac{1}{2}$$

or, $$\frac{ab+2ax-x^2}{a(a+b)} = \frac{1}{2}$$

or, $$2x^2-4ax+(a^2-ab) = 0.$$

$$\therefore\ x = \frac{4a \pm \sqrt{16a^2-8(a^2-ab)}}{4}$$

$$= a \pm \tfrac{1}{4}\sqrt{2a^2+2ab} = a \pm \sqrt{\frac{a(a+b)}{2}}.$$

Now, $a-\sqrt{\dfrac{a(a+b)}{2}} < a$

and $a-\sqrt{\dfrac{a(a+b)}{2}} > a-a=0$ if $b < a$.

$\therefore\ a-\sqrt{\dfrac{a(a+b)}{2}} \in (0, a)$.

Also $a+\sqrt{\dfrac{a(a+b)}{2}} \notin [0, a]$.

Hence, $x=a-\sqrt{\dfrac{a(a+b)}{2}}$ is the unique solution of

$$F(x)=\tfrac{1}{2} \text{ if } a > b.$$

So the required median is $a-\sqrt{\dfrac{a(a+b)}{2}}$.

Ex. 29. *If X be a normal (m, σ) variate, then prove that*

$$\mu_{2r+2}=\sigma^2\mu_{2r}+\sigma^3\frac{d\mu_{2r}}{d\sigma}.$$

Hence, find the coefficient of kurtosis β_2 of this distribution.

We have $\mu_{2r}=\dfrac{1}{\sqrt{2\pi}\,\sigma}\displaystyle\int_{-\infty}^{\infty}(x-m)^{2r}\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx.$

Assuming the validity of differentiation under the integral sign with respect to σ, we get

$$\frac{d\mu_{2r}}{d\sigma}=\frac{1}{\sqrt{2\pi}}\left\{-\frac{1}{\sigma^2}\int_{-\infty}^{\infty}(x-m)^{2r}\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx\right.$$

$$\left.+\frac{1}{\sigma}\int_{-\infty}^{\infty}(x-m)^{2r+2}\,\frac{1}{\sigma^3}\, e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx\right\}$$

or, $\dfrac{d\mu_{2r}}{d\sigma}=\dfrac{1}{\sqrt{2\pi}}\left(-\dfrac{\sqrt{2\pi}}{\sigma}\mu_{2r}+\dfrac{1}{\sigma^4}\sqrt{2\pi}\,\sigma\mu_{2r+2}\right)$

or, $\dfrac{d\mu_{2r}}{d\sigma}=-\dfrac{1}{\sigma}\mu_{2r}+\dfrac{1}{\sigma^3}\mu_{2r+2},$

i.e., $\mu_{2r+2}=\sigma^2\mu_{2r}+\sigma^3\dfrac{d\mu_{2r}}{d\sigma}.$

Then $\mu_4=\sigma^2\mu_2+\sigma^3\dfrac{d\mu_2}{d\sigma}$,

But $\mu_2=\sigma^2$ $\therefore$ $\mu_4=\sigma^4+2\sigma^4=3\sigma^4$.

So, $\beta_2=\dfrac{\mu_4}{\sigma^4}=3$.

Ex. 30. *Obtain the recurrence relation*

$$\mu_{k+1}=\mu\left(k\mu_{k-1}+\frac{d\mu_k}{d\mu}\right)$$

for the Poisson distribution with paraneter μ. *Hence find the coefficient of skewness and the coefficient of excess of the Poisson distribution.*

Here we note that the mean of the given distribution is μ.

We have $\mu_k=\sum_{x=0}^{\infty}(x-\mu)^k\dfrac{e^{-\mu}\mu^x}{x!}$,

where the infinite series in the right hand side is absolutely convergent.

$$\therefore\ \frac{d\mu_k}{d\mu}=\sum_{x=0}^{\infty}\frac{d}{d\mu}\left\{(x-\mu)^k\frac{e^{-\mu}\mu^x}{x!}\right\}$$

(Here the process of term by term differentiation is valid)

$$=\sum_{x=0}^{\infty}\left[-\frac{e^{-\mu}\mu^x}{x!}(x-\mu)^k-\frac{k(x-\mu)^{k-1}e^{-\mu}\mu^x}{x!}+\frac{x\mu^{x-1}e^{-\mu}(x-\mu)^k}{x!}\right]$$

$$=\sum_{x=0}^{\infty}\frac{(x-\mu)^k e^{-\mu}\mu^x}{x!}\left(-1+\frac{x}{\mu}\right)-k\,\mu_{k-1}$$

$$=\frac{1}{\mu}\sum_{x=0}^{\infty}\frac{(x-\mu)^{k+1}e^{-\mu}\mu^x}{x!}-k\,\mu_{k-1}$$

$$=\frac{\mu_{k+1}}{\mu}-k\,\mu_{k-1}.$$

Thus $\mu_{k+1}=\mu\left(k\,\mu_{k-1}+\dfrac{d\mu_k}{d\mu}\right)$.

Then $\mu_3=\mu\left(2\mu_1+\dfrac{d\mu_2}{d\mu}\right)$.

But $\mu_1=0$.

$\therefore$ $\mu_3=\mu\dfrac{d\mu_2}{d\mu}$. (7.8.4

Again, $\mu_3=\mu\left(\mu_0+\frac{d\mu_1}{d\mu}\right)$ and $\mu_0=1$.

$\therefore\quad \mu_2=\mu.$

Hence, from (7.7.4), $\mu_3=\mu \cdot 1=\mu$.

$$\therefore\quad \gamma_1=\frac{\mu_3}{\sigma^3}=\frac{\mu}{\mu\sqrt{\mu}}=\frac{1}{\sqrt{\mu}}.$$

$\left[\because\right.$ For poisson $\sigma=\sqrt{\text{var}X}=\sqrt{\mu}$
$\text{Var}X=\mu$

Again, $\mu_4=\mu\left(3\mu_2+\frac{d\mu_3}{d\mu}\right)$

$\therefore\quad \mu_4=3\mu^2+\mu \cdot 1=3\mu^2+\mu.$

$$\therefore\quad \gamma_2=\frac{\mu_4}{\sigma^4}-3=\frac{3\mu^2+\mu}{\mu^2}-3=\frac{1}{\mu}.$$

Ex. 31. *Find the mean and variance of the Maxwell distribution defined by the probability density function, given by,*

$$f(x)=\frac{4}{\beta^3\sqrt{\pi}}\,x^2e^{-\frac{x^2}{\beta^2}},\ \beta>0,\ \textit{if}\ \ 0<x<\infty$$

$$=0,\ \textit{elsewhere}.$$

The mean m is given by

$$m=\int_0^\infty \frac{4}{\beta^3\sqrt{\pi}}\,x^3e^{-\frac{x^2}{\beta^2}}\,dx,$$

(the integral is absolutely convergent)

$$=\frac{2\beta}{\sqrt{\pi}}\int_0^\infty ze^{-z}\,dz, \qquad \left(\frac{x^2}{\beta^2}=z\right)$$

$$=\frac{2\beta}{\sqrt{\pi}}\,\Gamma(2)=\frac{2\beta}{\sqrt{\pi}}.$$

If X be the corresponding random variable, then

$$E(X^2)=\frac{4}{\beta^3\sqrt{\pi}}\int_0^\infty x^4e^{-\frac{x^2}{\beta^2}}\,dx$$

$$=\frac{2\beta^2}{\sqrt{\pi}}\int_0^\infty z^{\frac{3}{2}}e^{-z}\,dz$$

$$=\frac{2\beta^2}{\sqrt{\pi}}\,\Gamma(\tfrac{5}{2})=\frac{3\beta^2}{2}.$$

$$\therefore\quad \text{Var}\,(X)=E(X^2)-m^2=\frac{3\beta^2}{2}-\frac{4\beta^2}{\pi}=(3\pi-8)\frac{\beta^2}{2\pi}.$$

Ex. 32. *X has uniform distribution in $(-a, a)$. Find the moment generating function of X and hence find the central moments of the distribution of X.*

The moment generating function $M_X(t)$ of X is given by

$$M_X(t)=\int_{-a}^{a}\frac{1}{2a}e^{tx}\,dx, \text{ where we note that } \int_{-a}^{a}\frac{1}{2a}e^{tx}\,dx$$

is absolutely convergent for all $t \in R$.

$$\text{Now } \int_{-a}^{a}\frac{1}{2a}e^{tx}\,dx=\frac{1}{2at}(e^{at}-e^{-at}) \text{ if } t\neq 0$$

$$=1 \text{ if } t=0$$

$$\therefore\ M_X(t)=\frac{1}{at}\sinh(at) \text{ if } t\neq 0$$

$$=1 \qquad \text{if } t=0.$$

$$\text{Now } \tfrac{1}{2}(e^{at}-e^{-at})=at+\frac{(at)^3}{3!}+\frac{(at)^5}{5!}+\cdots\cdots+, \text{ for all } t\in R.$$

$$\therefore\ \frac{1}{at}\sinh(at)=1+\frac{(at)^2}{3!}+\frac{(at)^4}{5!}+\cdots\cdots \text{ if } t\neq 0.$$

Further, we observe that $M_X(0)$, $M_X'(0)$, $M_X''(0)$, ...exist and $M_X(0)=1$. Hence, the expansion of $M_X(t)$ in Maclaurin's infinite series is given by

$$M_X(t)=1+\frac{(at)^2}{3!}+\frac{(at)^4}{5!}+\cdots\cdots \text{ for all } t\in R.$$

Then the kth order moment of X about the origin is given by

$$\alpha_k=\text{coefficient of } \frac{t^k}{k!} \text{ in the above expansion of } M_X(t)$$

$$=0 \text{ if } k \text{ is an odd positive integer}$$

$$=\frac{a^k}{(k+1)!}k! \text{ if } k \text{ is an even positive integer.}$$

So the mean m of the distribution of X is given by $m=\alpha_1=0$.

The kth order central moment μ_k is given by

$$\mu_k=E\{(X-0)^k\}=E(X^k)=\alpha_k.$$

$$\therefore\ \mu_k=0, \text{ if } k \text{ is an odd positive integer}$$

$$=\frac{a^k}{k+1}, \text{ if } k \text{ is an even positive integer.}$$

Ex. 33. *Find the moment generating function of the continuous random variable X with probability density function given by*

$$f(x)=\frac{2a+x}{4a^2} \quad if\ -2a \leqslant x \leqslant 0$$

$$=\frac{2a-x}{4a^2} \quad if\ 0 < x \leqslant 2a$$

$$=0, \quad elsewhere.$$

The moment generating function $M_X(t)$ is given by

$$M_X(t)=E(e^{tX})$$

$$=\int_{-2a}^{2a} e^{tx}\ f(x)\ dx,$$

provided the integral is absolutely convergent.

Here $\int_{-2a}^{2a} e^{tx}\ f(x)\ dx$ is a proper integral and it exists for all real values of t.

Now $\int_{-2a}^{2a} e^{tx}\ f(x)\ dx$

$$=\frac{1}{4a^2}\int_{-2a}^{0} e^{tx}\ (2a+x)\ dx+\frac{1}{4a^2}\int_{0}^{2a} e^{tx}\ (2a-x)\ dx$$

$$=\frac{1}{2a}\int_{-2a}^{2a} e^{tx}\ dx+\frac{1}{4a^2}\int_{-2a}^{0} xe^{tx}dx-\frac{1}{4a^2}\int_{0}^{2a} xe^{tx}\ dx$$

$$=\frac{1}{2a}\cdot\frac{e^{2at}-e^{-2at}}{t}-\frac{1}{4a^2}\int_{0}^{2a} ue^{-tu}\ du$$ → $u = x$ put

$$-\frac{1}{4a^2}\int_{0}^{2a} xe^{tx}\ dx \quad \text{if } t \neq 0$$

$$=\frac{\sinh(2at)}{at}-\frac{1}{2a^2}\int_{0}^{2a} x\cosh(tx)\ dx$$

$$=\frac{\sinh(2at)}{at}-\frac{\sinh(2at)}{at}+\frac{\cosh(2at)}{2a^2t^2}-\frac{1}{2a^2t^2}$$

$$=\frac{-1+\cosh(2at)}{2a^2t^2}$$

$$=\frac{-1+\cosh^2(at)+\sinh^2(at)}{2a^2t^2}$$

$$=\frac{2\sinh^2(at)}{2a^2t^2}$$

$$=\left\{\frac{\sinh(at)}{at}\right\}^2 \quad \text{if } t\neq 0.$$

Also $M_X(0)=1$.

$$\therefore \quad M_X(t)=\left[\frac{\sinh(at)}{at}\right]^2 \quad \text{if } t\neq 0$$

$$=1 \qquad \text{if } t=0.$$

2010

Ex. 34. *Find the kth order moment about the origin of the continuous random variable X with probability density function* $f(x)=2e^{-2x}$, $0<x<\infty$, *by finding the moment generating function of X.*

Most alike

The moment generating function $M_X(t)$ is given by

$$M_X(t)=\int_0^\infty 2e^{tx}\,e^{-2x}\,dx=2\int_0^\infty e^{(t-2)x}\,d$$

$$=\frac{2}{2-t} \quad \text{if } t<2.$$

Now $\dfrac{2}{2-t}=\left(1-\dfrac{t}{2}\right)^{-1}=1+\dfrac{t}{2}+\left(\dfrac{t}{2}\right)^2+\ldots\ldots$ if $\left|\dfrac{t}{2}\right|<1$.

$\therefore$ $\alpha_k=$ coefficient of $\dfrac{t^k}{k!}$ in the above expansion

$$=\frac{k!}{2^k}.$$

So, the required kth order moment is $\dfrac{k!}{2^k}$.

Ex. 35. *Find the characteristic function of the Laplace distribution defined by the probability density function $f(x)$, given by,*

$$f(x)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}}, \quad -\infty < x < \infty \ (\lambda > 0).$$

The required characteristic function $\phi(t)$ is given by

$$\phi(t)=\int_{-\infty}^{\infty} e^{itx}\cdot\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}}\,dx$$

$$=\frac{1}{2\lambda}\int_{-\infty}^{\mu} e^{itx}\,e^{\frac{x-\mu}{\lambda}}\,dx+\frac{1}{2\lambda}\int_{\mu}^{\infty} e^{itx}\,e^{\frac{-x+\mu}{\lambda}}\,dx$$

$$=\frac{1}{2\lambda}e^{\frac{-\mu}{\lambda}}\int_{-\infty}^{\mu} e^{\left(it+\frac{1}{\lambda}\right)x}\,dx+\frac{1}{2\lambda}e^{\frac{\mu}{\lambda}}\int_{\mu}^{\infty} e^{-\left(-it+\frac{1}{\lambda}\right)x}\,dx$$

$$=\frac{1}{2\lambda}e^{-\frac{\mu}{\lambda}}\frac{e^{\left(it+\frac{1}{\lambda}\right)\mu}}{it+\frac{1}{\lambda}}-\frac{1}{2\lambda}e^{\frac{\mu}{\lambda}}\frac{e^{\left(it-\frac{1}{\lambda}\right)\mu}}{it-\frac{1}{\mu}}$$

$$=\frac{1}{2}\left(\frac{e^{it\mu}}{1+it\lambda}-\frac{e^{it\mu}}{it\lambda-1}\right)$$

$$=\frac{e^{it\mu}}{1+t^2\lambda^4}.$$

So the characteristic function is $\dfrac{e^{it\mu}}{1+\lambda^2 t^2}$ defined for all real values of t.

Ex. 35. *For a given positive integer n, show that the standard deviation of a binomial (n, p) distribution cannot exceed $\frac{\sqrt{n}}{2}$. For a given n, find also the coefficient of skewness of the binomial distribution with maximum standard deviation.*

Let X be a binomial (n, p) variate.

Then $\text{Var}(X)=np(1-p)$. Hence the standard deviation σ of X is given by $\sigma=\sqrt{n}\sqrt{p(1-p)}$.

Now $p > 0$ and $1-p > 0$.

$\therefore \quad \sqrt{p(1-p)} \leqslant \frac{p+(1-p)}{2}$ (taking A.M. and G.M. of $p, 1-p$),

i.e., $\sqrt{p(1-p)} \leqslant \frac{1}{2}$,

where the equality sign occurs if $p=1-p$ *i.e.*, if $p=\frac{1}{2}$.

Hence, $\sigma \leqslant \sqrt{n} \cdot \frac{1}{2}$ and the sign of equality occurs if $p=\frac{1}{2}$. Thus it is proved that the standard deviation of a binomial (n, p) variate cannot exceed $\frac{\sqrt{n}}{2}$ and further the standard deviation attains the maximum value $\frac{\sqrt{n}}{2}$ if $p=\frac{1}{2}$. Thus the required co-efficient of skewness is

$$\gamma_1=\frac{1-2p}{\sqrt{np(1-p)}}=0.$$

Ex. 37. *Find the probability density function of the continuous distribution which has the characteristic function* $e^{-|t|}$

Let X be the continuous random variable whose characteristic function is $e^{-|t|}$. Then by the inversion formula (7.6.7).

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\, e^{-|t|}\, dt,$$

where $f(x)$ is the probability density function of X.

$$=\frac{1}{2\pi}\int_{-\infty}^{0} e^{+t}\, e^{-itx}\, dt+\frac{1}{2\pi}\int_{0}^{\infty} e^{-t}\, e^{-itx}\, dt$$

$$=\frac{1}{2\pi}\int_{0}^{\infty} e^{-t}\, e^{itx}\, dt+\frac{1}{2\pi}\int_{0}^{\infty} e^{-t}\, e^{-itx}\, dt$$

$$=\frac{1}{2\pi}\int_{0}^{\infty} e^{-t}(e^{itx}+e^{-itx})\, dt$$

$$=\frac{1}{\pi}\int_{0}^{\infty} e^{-t} \cos tx\, dt.$$

$$= \frac{1}{\pi} \underset{B \to \infty}{Lt} \int_0^B e^{-t} \cos tx \, dt$$

$$= \frac{1}{\pi} \underset{B \to \infty}{Lt} \left\{ \frac{e^{-t}(-\cos tx + x \sin tx)}{1+x^2} \right\}_0^B$$

$$= \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

Ex. 38. *Prove that for any real characteristic function $\phi(t)$, the inequality $1-\phi(2t) \leqslant 4\{1-\phi(t)\}$ holds and hence prove that for any characteristic function $1-|f(2t)|^2 \leqslant 4\{1-|f(t)|^2\}$.*

Let X be the random variable whose characteristic function is $\phi(t)$. Then $\phi(t)=E(e^{itX})$. Since $\phi(t)$ is real, we have $\phi(t)=\overline{\phi(t)}$ and consequently

$$\phi(t) = \tfrac{1}{2}\{\phi(t)+\overline{\phi(t)}\}$$
$$= \tfrac{1}{2}E(2 \cos tX)$$
$$= E(\cos tX).$$

$$\therefore \quad 1-\phi(2t) = 1-E(\cos 2tX)$$
$$= E(1-\cos 2tX)$$
$$= 2E(\sin^2 tX).$$

Now $\sin^2 tX = 4 \sin^2 \dfrac{tX}{2} \cos^2 \dfrac{tX}{2} \leqslant 4 \sin^2 \dfrac{tX}{2}.$

So, $2E(\sin^2 tX) \leqslant 8E\left(\sin^2 \dfrac{tX}{2}\right) = 4E(1-\cos tX).$

$\therefore \; 1-\phi(2t) \leqslant 4\{1-\phi(t)\}.$

Now for any characteristic function $f(t)$,

$$f(t)\overline{f(t)} = |f(t)|^2,$$

is a real valued continuous function of t. So by the uniqueness theorem, there exists a distribution whose characteristic function is $|f(t)|^2$ which is real.

Let $\phi(t)=|f(t)|^2$. Then by the above inequality for real characteristic function, we have

$$1-\phi(2t) \leqslant 4\{1-\phi(t)\}.$$

But we have $\phi(2t)=|f(2t)|^2$ and $\phi(t)=|f(t)|^2$.

Hence we get $1-|f(2t)|^2 \leqslant 4\{1-|f(t)|^2\}$.

Ex. 39. *Find the characteristic function of the Pascal distribution with probability mass function given by*

$$P(X=r)=\frac{1}{1+m}\left(\frac{m}{1+m}\right)^r, \ (m>0) \text{ for } r=0, 1, 2, \ldots\ldots$$

Hence, find the mean and variance of X.

If $\phi(t)$ be the characteristic function of X, then

$$\phi(t)=\sum_{r=0}^{\infty}\frac{1}{1+m}\left(\frac{m}{1+m}\right)^r e^{itr}.$$

Now $\displaystyle\sum_{r=0}^{\infty}\frac{1}{1+m}\left(\frac{m}{1+m}\right)^r e^{itr}.$

$$=\frac{1}{1+m}\sum_{r=0}^{\infty}\left(\frac{m}{1+m}e^{it}\right)^r$$

$$=\frac{1}{1+m}\frac{1}{1-\frac{m}{1+m}e^{it}} \quad \left(\because \left|\frac{m}{1+m}e^{it}\right|=\frac{m}{1+m}<1 \text{ for } m>0\right)$$

$$=\frac{1}{1-m(e^{it}-1)}, \quad \text{for all } t\in R.$$

Now, $\displaystyle\phi(t)=\frac{1}{1+m}\left(1-\frac{m}{1+m}e^{it}\right)^{-1}$

$$=\frac{1}{1+m}\left\{1+\frac{me^{it}}{1+m}+\frac{m^2}{(1+m)^2}e^{2it}+\cdots\right\}.$$

We see that the coefficient of it in the above expansion of $\phi(t)$ is

$$\frac{1}{1+m}\left\{\frac{m}{1+m}+\frac{m^2}{(1+m)^2}\cdot 2+\frac{m^3}{(1+m)^3}\cdot 3+\cdots\right\}$$

$$=\frac{m}{(1+m)^2}\left\{1+\frac{2m}{1+m}+3\left(\frac{m}{1+m}\right)^2+\cdots\right\}$$

$$=\frac{m}{(1+m)^2}\left(1-\frac{m}{1+m}\right)^{-2}, \quad \because \left|\frac{m}{1+m}\right|<1 \text{ if } m>0$$

$$=m.$$

Hence the mean of X is m.

Again, α_2 = the coefficient of $\dfrac{(it)^2}{2!}$ in the expansion of $\phi(t)$

$$=\frac{1}{1+m}\left\{\frac{m}{1+m}+\frac{m^2}{(1+m)^2}\cdot 2^2+\frac{m^3}{(1+m)^3}3^2+\cdots\right\}$$

$$=\frac{m}{(1+m)^2}\left\{1+\frac{m}{1+m}\cdot 2^2+\left(\frac{m}{1+m}\right)^2\cdot 3^2+\cdots\right\}$$

Now, we have

$$(1-x)^{-2}=1+2x+3x^2+\cdots\cdots,\ \text{if}\ |x|<1.$$

$$\therefore\ \frac{x}{(1-x)^2}=x+2x^2+3x^3+4x^4+\cdots\cdots,\ \text{if}\ |x|<1.$$

Now a power series can be differentiated term by term in the interval of convergence.

$$\therefore\ \frac{d}{dx}\left[\frac{x}{(1-x)^2}\right]=1+2^2x+3^2x^2+4^2x^3+\cdots,\ \text{if}\ |x|<1$$

or, $$\frac{1}{(1-x)^2}+\frac{2x}{(1-x)^3}=1+2^2x+3^2x^2+4^2x^3+\cdots,\ \text{if}\ |x|<1.$$

$$\therefore\ 1+2^2x+3^2x^2+4^2x^3+\cdots=\frac{1+x}{(1-x)^3},\ \text{if}\ |x|<1.$$

Now, $0<\dfrac{m}{1+m}<1.$

So we have,

$$1+2^2\frac{m}{1+m}+3^2\left(\frac{m}{1+m}\right)^2+\cdots$$

$$=\frac{1+\dfrac{m}{1+m}}{\left(1-\dfrac{m}{1+m}\right)^3}=(1+2m)(1+m)^2.$$

Hence, $\alpha_2=m(1+2m).$

$$\therefore\ \text{Var}(X)=\alpha_2-m^2$$
$$=m(1+2m)-m^2$$
$$=m(m+1).$$

Ex. 40. *Applying 7.4.12 and 7.4.18, find (i) the mean and variance of binomial* $(2, \frac{1}{2})$ *distribution, (ii) the mean and variance of the random variable U defined in example 3 of 7.2.*

[C. H. (Math.) '90]

(i) If X be the binomial $(2, \frac{1}{2})$ variate, then the corresponding distribution function $F_X(x)$ is given by

$$\begin{aligned} F_X(x) &= 0 \text{ if } -\infty < x < 0 \\ &= {}^2C_0(\tfrac{1}{2})^2 \text{ if } 0 \leqslant x < 1 \\ &= {}^2C_0(\tfrac{1}{2})^2 + {}^2C_1(\tfrac{1}{2})^2 \text{ if } 1 \leqslant x < 2 \\ &= 1 \text{ if } x \geqslant 2. \end{aligned}$$

Then by 7.4.12 we have

$$E(X) = \int_0^\infty \{1 - F_X(x) - F_X(-x)\}\, dx$$

$$\int_0^\infty \{1 - F_X(x)\}\, dx$$

[$\because$ here $F_X(-x) = 0$ for $x > 0$.]

$$= \int_0^1 (1 - \tfrac{1}{4})\, dx + \int_1^2 (1 - \tfrac{3}{4})\, dx$$

$$= \tfrac{3}{4} + \tfrac{1}{4} = 1.$$

Now by 7.4.18, we have

$$\text{Var}(X) = \int_0^\infty 2x\{1 - F_X(x) + F_X(-x)\}\, dx - \{E(X)\}^2$$

$$= \int_0^\infty 2x\{1 - F_X(x)\}\, dx - 1$$

$$= \int_0^1 2x(1 - \tfrac{1}{4})\, dx + \int_1^2 2x(1 - \tfrac{3}{4})\, dx - 1$$

$$= \tfrac{3}{4} + \tfrac{3}{4} - 1$$

$$= \tfrac{1}{2}.$$

(ii) By example 3 of 7.2, the distribution function $F_U(u)$ of U is given by

$$F_U(u)=0 \quad \text{if } u \leqslant -2$$
$$=\frac{u+2}{4} \quad \text{if } -2<u<1$$
$$=1 \quad \text{if } u \geqslant 1.$$

Now by 7.4.12,

$$E(U)=\int_0^\infty \{1-F_U(u)-F_U(-u)\}\,du$$

$$=\int_0^\infty \{1-F_U(u)\}\,du-\int_0^\infty F_U(-u)\,du$$

$$=\int_0^1 \left(1-\frac{u+2}{4}\right)du+\int_1^\infty (1-1)\,du-\int_0^2 F_U(-u)\,du-\int_2^\infty F_U(-u)\,du$$

$$=\int_0^1 \frac{2-u}{4}\,du-\int_0^2 F_U(-u)\,du$$

$$=\int_0^1 \frac{2-u}{4}\,du-\int_0^2 \frac{-u+2}{4}\,du$$

$$=\tfrac{3}{8}-\tfrac{1}{2}$$

$$=-\tfrac{1}{8}.$$

Now, $\text{Var}(U)=\int_0^\infty 2u\{1-F_U(u)+F_U(-u)\}\,du-\{E(U)\}^2$

$$=2\int_0^\infty u\{1-F_U(u)\}\,du+2\int_0^\infty u\,F_U(-u)\,du-\tfrac{1}{64}$$

$$=2\int_0^1 u\left\{1-\frac{u+2}{4}\right\}du+2\int_1^\infty u(1-1)\,du+2\int_0^2 uF_U(-u)du$$

$$+2\int_2^\infty u\,F_U(-u)-\tfrac{1}{64}$$

$$=\tfrac{1}{2}\int_0^1 u(2-u)\,du+\tfrac{1}{2}\int_0^2 u(-u+2)\,du+0-\tfrac{1}{64}$$

$$=\tfrac{1}{2}(1-\tfrac{1}{3})+\tfrac{1}{2}(-\tfrac{8}{3}+4)-\tfrac{1}{64}$$

$$=\tfrac{63}{64}.$$

Ex. 41. *Find the mean and variance of the random variable whose distribution function is*

$$F(x)=1-pe^{-\lambda x} \quad for \ 0 \leqslant x < \infty$$
$$=0, \quad elsewhere,$$
$$(\lambda > 0, \ p > 0).$$

$$E(X)=\int_0^{\infty} \{1-F(x)-F(-x)\}\,dx$$

$$=\int_0^{\infty} \{1-F(x)\}\,dx$$

$$=p\int_0^{\infty} e^{-\lambda x}dx$$

$$=\frac{p}{\lambda}.$$

$$\text{Var}\,(X)=\int_0^{\infty} 2x\{1-F(x)+F(-x)\}\,dx-\{E(X)\}^2$$

$$=2\int_0^{\infty} x\left(pe^{-\lambda x}\right)dx-\frac{p^2}{\lambda^2}$$

$$=\frac{2p}{\lambda^2}\int_0^{\infty} ue^{-u}\,du-\frac{p^2}{\lambda^2}$$

$$=\frac{2p}{\lambda^2}\,\Gamma(2)-\frac{p^2}{\lambda^2}$$

$$=\frac{2p}{\lambda^2}-\frac{p^2}{\lambda^2}$$

$$=\frac{p(2-p)}{\lambda^2}.$$

Ex. 42. *X is a discrete random variable with probability mass function given by*

$$P(X=-2)=P(X=0)=\tfrac{1}{4},$$
$$P(X=1)=\tfrac{1}{3},\ P(X=2)=\tfrac{1}{6}.$$

Find the median of the distribution of X.

The distribution function $F(x)$ of X is given by

$$F(x)=0 \text{ if } -\infty < x < -2$$
$$=\tfrac{1}{4} \text{ if } -2 \leqslant x < 0$$
$$=\tfrac{1}{4}+\tfrac{1}{4} \text{ if } 0 \leqslant x < 1$$
$$=\tfrac{1}{4}+\tfrac{1}{4}+\tfrac{1}{3} \text{ if } 1 \leqslant x < 2$$
$$=1 \text{ if } x \geqslant 2.$$

Here we see that $F(x)=\frac{1}{2}$ for all x satisfying $0 \leqslant x < 1$ and $F(x) > \frac{1}{2}$ for all $x \geqslant 1$. Then $F(x-0)=\frac{1}{2}$ for all $x \in [0, 1]$ and $F(x) \geqslant \frac{1}{2}$ for all $x \in [0, 1]$. So we can say that $F(x-0) \leqslant \frac{1}{2}$ and $F(x) \geqslant \frac{1}{2}$ for all $x \in [0, 1]$ and these inequalities are not satisfied simultaneously by any other value of x and hence, by definition the required median is $\dfrac{0+1}{2}=\dfrac{1}{2}$.

Ex. 43. *Find the value of l (if there be any) for which the distribution of a $\gamma(l)$ variate is mesokurtic.*

Let X be a $\gamma(l)$ variate. We know that $E(X)=l$, Var $(X)=l$, $\mu_4=3l^2+6l$, by (gamma distribution) *Page.* 393.

Then X will have mesokurtic distribution if

$$\frac{\mu_4}{\sigma^4}=3, \quad \text{i.e., if } \frac{3l^2+6l}{l^2}=3,$$

which gives $6=0$ and this is absurd.

Hence, there exists no value of l for which X has mesokurtic distribution.

Ex. 44. *If X is a normal $(0, \sigma)$ variate, then find the variance of $X+X^2$.*

Here $E(X)=0$, $E(X^2)=E\{(X-0)^2\}=\text{Var }(X)=\sigma^2$.

So, $E(X+X^2)=\sigma^2$.

Now, $\text{Var }(X+X^2)=E\{(X+X^2)^2\}-\{E(X+X^2)\}^2$

$$=E(X^2)+E(X^4)+2E(X^3)-\sigma^4$$
$$=\sigma^2+E(X^4)+2E(X^3)-\sigma^4.$$

Now, since X is a normal $(0, \sigma)$ variate,

$$E(X^4)=\mu_4=1.3.\sigma^4=3\sigma^4 \text{ and } E(X^3)=0.$$

$$\therefore \text{Var }(X+X^2)=\sigma^2+3\sigma^4-\sigma^4$$
$$=\sigma^2(1+2\sigma^2).$$

Ex. 45. *Find the mode and the median of the continuous distribution with probability density function given by*

$$f(x)=\frac{abx^{a-1}}{(1+bx^a)^2},\ b>0,\ a>1,\ 0<x<\infty.$$

Here $f'(x)=\dfrac{ab(a-1)x^{a-2}}{(1+bx^a)^2}-\dfrac{2abx^{a-1}\cdot bax^{a-1}}{(1+bx^a)^3},\ 0<x<\infty.$

Hence, $f'(x)=0$ gives $(a-1)-\dfrac{2abx^a}{bx^a+1}=0$

or, $x^a=\dfrac{a-1}{b(a+1)}$,

i.e., $x=\left\{\dfrac{a-1}{b(a+1)}\right\}^{\frac{1}{a}}=C$ (say).

It can be shown that $f''(C)<0$. So $f(x)$ has a local maximum at $x=C$. Hence, the mode is $\left\{\dfrac{a-1}{b(a+1)}\right\}^{\frac{1}{a}}$.

The corresponding distribution function $F(x)$ is given by

$$F(x)=\int_0^x \frac{abt^{a-1}}{(1+bt^a)^2}\,dt,\ x>0.$$

$$=0,\qquad \text{elsewhere},$$

i.e., $F(x)=1-\dfrac{1}{1+bx^a}$, if $x>0$

$=0$, elsewhere.

Then $F(x)=\frac{1}{2}(x>0)$ gives $\dfrac{1}{1+bx^a}=\dfrac{1}{2}$

or, $x=\left(\dfrac{1}{b}\right)^{\frac{1}{a}}$.

Hence, the required median is $\left(\dfrac{1}{b}\right)^{\frac{1}{a}}$.

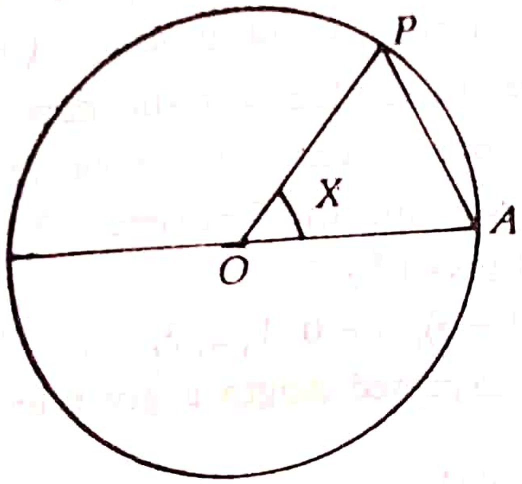
Examples VII

1. Find the mean and variance of each of the continuous distributions given by the following probability density functions :

(i) $f(x)=\frac{1}{2}-ax$ if $0 < x < 4$ (ii) $f(x)=\frac{1}{2}e^{-|x|}$, $-\infty<x<\infty$
$=0$, elsewhere.

2. A point is selected at random on a circle of unit radius. Compute the mathematical expectation of its distance from a fixed point of the circle. [*C. H. (Math.) '84*]

[*Hint :* O is the centre of the circle of unit radius and X is the random variable denoting the measure of $\angle AOP$, where A is a fixed point on the circle and P is the point selected at random on



the circle. Here X is uniform in $(0, 2\pi)$. The random variable denoting the distance AP is $\left|2 \sin \frac{X}{2}\right|=2 \sin \frac{X}{2}$. The required expectation is $E\left(2 \sin \frac{X}{2}\right)=\int_0^{2\pi}\left(2 \sin \frac{x}{2}\right)\frac{1}{2\pi}dx=\frac{4}{\pi}$.]

3. If 4 balls are drawn (a) with replacement or (b) without replacement from an urn containing 8 white and 3 black balls, then find the expectation of the number of white balls in the cases (a) and (b).

[*Hint :* Let X be the random variable denoting the number of white balls drawn.

(a) X is binomial $(4, p)$ where p=probability of drawing a white ball$=\frac{8}{11}$, and hence, $E(X)=4p=\frac{32}{11}$.

(b) $$P(X=1)=\frac{{}^8C_1 \times {}^3C_3}{{}^{11}C_4}=\frac{8}{{}^{11}C_4}$$

$$P(X=2)=\frac{{}^8C_2 \times {}^3C_2}{{}^{11}C_4}=\frac{84}{{}^{11}C_4}$$

$$P(X=3)=\frac{{}^8C_3 \times {}^3C_1}{{}^{11}C_4}=\frac{168}{{}^{11}C_4}$$

$$P(X=4)=\frac{{}^8C_4}{{}^{11}C_4}=\frac{70}{{}^{11}C_4}.$$

$$\therefore \quad E(X)=1 \cdot \frac{8}{{}^{11}C_4}+2 \cdot \frac{84}{{}^{11}C_4}+3 \cdot \frac{168}{{}^{11}C_4}+4 \cdot \frac{70}{{}^{11}C_4}=\frac{32}{11}.\Big]$$

4. A coin is tossed repeatedly and the probability that a head appears at any toss is p, where $0 < p=1-q < 1$. Find the expected length of the initial run of heads. [*C. H. (Math.)* '90]

[*Hint* : Let X be the random variable denoting the length of the initial run of heads. Then X denotes the number of heads before the tail appears for the first time. Then the probability mass function of X is given by

$$P(X=r)=p^r(1-p), \quad r=0, 1, 2, 3, \ldots\ldots$$

$\therefore$ the required expected length is given by

$$E(X)=\sum_{r=0}^{\infty} rp^r(1-p)$$

$$=p(1-p)(1+2p+3p^2+\cdots)$$

$$=p(1-p)(1-p)^{-2}=\frac{p}{1-p}. \;]$$

5. The probability mass function of a discrete random variable X is given by

$$P(X=x)=kx, \quad x=1, 2, 3, \ldots, n.$$

Find $E(X)$.

6. Find the mean, variance, coefficient of skewness and coefficient of kurtosis of the distribution of a discrete random variable X, where the probability mass function of X is given by

$$p(X=0)=1-p, \qquad P(X=1)=p.$$

13. Find the mean and the variance of the continuous variate X with probability density function given by

$$f(x)=\tfrac{1}{2}\left(1-\frac{|x-1|}{2}\right), \quad -1<x<3$$
$$=0 \qquad , \text{ elsewhere.}$$

14. Find the mean and the median of the continuous random variable X with probability density function $f(x)$ given by

$$f(x)=2x, \quad 0 \leqslant x \leqslant 1$$
$$=0, \quad \text{elsewhere.}$$

15. The probability density function of a continuous distribution is given by

$$f(x)=a(x-x^2), \; 0 \leqslant x \leqslant 1$$

a being a constant. Find the mean, the mode and the median of this distribution.

16. Show that the variance of $\beta_2(m, n)$ variate is $\dfrac{m(m+n-1)}{(n-1)^2(n-2)}$ where $m>0, n>2$. [See 7.4.68]

17. Find the median of the distribution of X whose distribution function is given by

$$F(x)=0, \qquad x \leqslant 1$$
$$=\tfrac{1}{16}(x-1)^4, \quad 1<x \leqslant 3$$
$$=1, \qquad x>3.$$

Also find the mean of the distribution.

[*Hint* : $E(X)=\displaystyle\int_0^\infty \{1-F(x)-F(-x)\}\,dx=\int_0^\infty \{1-F(x)\}\,dx$

$$=\int_0^1 (1-0)\,dx+\int_1^3 \{1-\tfrac{1}{16}(x-1)^4\}\,dx+\int_3^\infty (1-1)\,dx=\tfrac{13}{5}.$$

Median $=1+\sqrt[4]{8}$.]

18. Pareto's distribution is a continuous distribution defined by the density function

$$f(x)=\frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} \quad \text{if } x>\beta$$
$$=0 \qquad \text{if } x \leqslant \beta$$

(where $\alpha>0, \beta>0$).

Now, $\left(x-\frac{a+b}{2}\right)^2-\left(\frac{b-a}{2}\right)^2=(x-a)(x-b)\leqslant 0$

$\therefore \left(x-\frac{a+b}{2}\right)^2 \leqslant \left(\frac{b-a}{2}\right)^2$. $\because a\leqslant x\leqslant b$.

$$\therefore E\left\{\left(X-\frac{a+b}{2}\right)^2\right\}=\int_a^b \left(x-\frac{a+b}{2}\right)^2 f(x)\,dx$$

$$\leqslant \left(\frac{b-a}{2}\right)^2 \int_a^b f(x)\,dx=\left(\frac{b-a}{2}\right)^2$$

$\therefore$ Var $(X) \leqslant \left(\frac{b-a}{2}\right)^2$.]

9. $F(x)$ denotes the distribution function of a continuous random variable X. Show that the expectation of X can be expressed as $E(X)=\int_0^\infty \{1-F(x)-F(-x)\}\,dx$. [C. H. (Math.) '94]

[See Theorem 7.4.3.]

10. The probability density function of a continuous random variable X is given by

$$\begin{aligned} f(x) &= \tfrac{1}{16}(3+x)^2, \quad -3\leqslant x\leqslant -1 \\ &= \tfrac{1}{16}(6-2x^2), \quad -1<x\leqslant 1 \\ &= \tfrac{1}{16}(3-x)^2, \quad 1<x\leqslant 3 \\ &= 0 \quad , \text{elsewhere.} \end{aligned}$$

Show that $E(X)=0$.

11. Find the mean and the standard deviation of the continuous distribution with probability density function given by

$$f(x)=ae^{-\frac{x}{b}},\ 0<x<\infty,\ b>0,\ a>0$$

and a, b are constants. Also find the semi-interquartile range of the distribution.

12. Find the mean deviation about the mean of the uniform distribution with probability density function given by

$$\begin{aligned} f(x) &= 1 \text{ if } 0<x<1 \\ &= 0, \text{ elsewhere.} \end{aligned}$$

Ex. VII

13. Find the mean and the variance of the continuous variate X with probability density function given by

$$f(x)=\tfrac{1}{2}\left(1-\frac{|x-1|}{2}\right), \quad -1<x<3$$
$$=0 \quad , \text{ elsewhere.}$$

14. Find the mean and the median of the continuous random variable X with probability density function $f(x)$ given by

$$f(x)=2x, \quad 0 \leqslant x \leqslant 1$$
$$=0, \quad \text{elsewhere.}$$

15. The probability density function of a continuous distribution is given by

$$f(x)=a(x-x^2), \; 0 \leqslant x \leqslant 1$$

a being a constant. Find the mean, the mode and the median of this distribution.

16. Show that the variance of $\beta_2(m, n)$ variate is $\dfrac{m(m+n-1)}{(n-1)^2(n-2)}$

where $m>0, n>2$. [See 7.4.68]

17. Find the median of the distribution of X whose distribution function is given by

$$F(x)=0, \quad x \leqslant 1$$
$$=\tfrac{1}{16}(x-1)^4, \quad 1<x \leqslant 3$$
$$=1, \quad x>3.$$

Also find the mean of the distribution.

[*Hint* : $E(X)=\displaystyle\int_0^\infty \{1-F(x)-F(-x)\}\,dx=\int_0^\infty \{1-F(x)\}\,dx$

$$=\int_0^1 (1-0)\,dx+\int_1^3 \{1-\tfrac{1}{16}(x-1)^4\}\,dx+\int_3^\infty (1-1)\,dx=\tfrac{13}{5}.$$

Median $=1+\sqrt[4]{8}$.]

18. Pareto's distribution is a continuous distribution defined by the density function

$$f(x)=\frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1} \quad \text{if } x>\beta$$
$$=0 \quad \text{if } x \leqslant \beta.$$

(where $\alpha>0, \beta>0$).

Show that the rth order moment about the origin exists if $\alpha > r$.
Show further that the variance of the distribution is

$$\frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)} \quad \text{if } \alpha > 2.$$

19. Find the mode (or modes) for a binomial $(n, \frac{1}{2})$ distribution.

20. Find the mode (or modes) and the coefficient of skewness γ_1 of a gamma distribution with parameter l.

[See 7.4.57].

*** 21. Prove that the recurrence relation

Puruhoz jee (P-200) (35)

$$\mu_{k+1} = p(1-p)\left(nk\,\mu_{k-1} + \frac{d\mu_k}{dp}\right)$$

for a binomial (n, p) variate. Hence obtain the expressions for β_1 and β_2.

22. If X is a $\gamma(l)$ variate, then find the value of $E(X^{\frac{3}{2}})$.

23. The distribution function $F(x)$ of a random variable X is given by

$$F(x) = 0 \qquad \text{if } x < 0$$
$$= 1 - \tfrac{1}{4}e^{-x} \text{ for } x \geqslant 0.$$

Find the mean and variance of X.

[*Hint* : Here X is neither discrete nor continuous.

$$m = E(X) = \int_0^\infty \{1 - F(x) - F(-x)\}\,dx$$

$$= \int_0^\infty \{1 - F(x)\}\,dx = \int_0^\infty \tfrac{1}{4}e^{-x}\,dx = \tfrac{1}{4}.$$

$$\text{Var}(X) = \int_0^\infty 2x\{1 - F(x) - F(-x)\}\,dx - m^2$$

$$= 2\int_0^\infty x\{1 - F(x)\}\,dx - \tfrac{1}{16} = 2\int_0^\infty x(\tfrac{1}{4}e^{-x})\,dx - \tfrac{1}{16}$$

$$= \tfrac{1}{2}\Gamma(2) - \tfrac{1}{16} = \tfrac{1}{2} - \tfrac{1}{16} = \tfrac{7}{16}.]$$

Ex. VII

24. Find the mode or modes (if exist) and the median of a continuous distribution with probability density function given by

$$f(x)=\lambda e^{-\lambda x}, x>0 \ (\lambda>0). \qquad [\textit{C. H. (Math.)} \text{ '82, '79}]$$

[*Hint :* Here $f'(x)$ exists for all $x>0$ and $f'(x)=-\lambda^2 e^{-\lambda x}$ for $x>0$, which shows that $f'(x)\neq 0$ for all $x>0$. So, $f(x)$ has no local maximum for any $x>0$. Again, $f(x)=0$ for $x\leqslant 0$ and so $f(x)$ has no local maximum for any $x<0$. Further, $f(0)=0$ and $f(x)>0$ for $x>0$ and so there exists no interval $(-\delta, \delta)$ $(\delta>0)$ such that $f(x)<f(0)$ for all $x\in(-\delta, \delta)$ and $x\neq 0$. So, $f(x)$ has no local maximum at $x=0$. Hence, the given distribution has no mode.]

25. An electronic device has a life span X (in units of 1000 hours), which is a continuous random variable with the probability density function

$$f(x)=e^{-x}, \quad x>0$$
$$=0, \quad \text{elsewhere.}$$

If the cost of one such item is Re. 1 and the manufacturer sells each item for Rs. 3, but guarantees a total refund if $X\leqslant 0{\cdot}8$, then find the manufacturer's expected profit per item.

[*Hint :* Let Z be the random variable denoting the profit. Then the spectrum of Z is $\{-1, 2\}$.

$$P(Z=-1)=P(X\leqslant 0{\cdot}8)=\int_0^{0\cdot 8} e^{-x}dx=1-e^{-0\cdot 8}$$

and $\quad P(Z=2)=P(X>0{\cdot}8)=e^{-0\cdot 8}.$

Then $E(Z)=(-1)(1-e^{-0\cdot 8})+2\,.\,e^{-0\cdot 8}=3e^{-0\cdot 8}-1.$]

26. Two players A and B throw with one die for a stake of Rs. 50, which is to be won by the player who first throws 6. If A starts, then find the expectation of A.

[*Hint* : Here the probability that A wins is

$$=\frac{1}{6}+\left(\frac{5}{6}\right)^2\cdot\frac{1}{6}+\left(\frac{5}{6}\right)^4\cdot\frac{1}{6}+\ldots\ldots$$

$$=\frac{1}{6}\left\{1-\left(\frac{5}{6}\right)^2\right\}^{-1}=\frac{6}{11}$$

and the probability that B wins is

$$\frac{5}{6}\cdot\frac{1}{6}+\left(\frac{5}{6}\right)^3\cdot\frac{1}{6}+\left(\frac{5}{6}\right)^5\cdot\frac{1}{6}+\ldots\ldots$$

$$=\frac{1}{6}\cdot\frac{5}{6}\left\{1-\left(\frac{5}{6}\right)^2\right\}^{-1}=\frac{5}{11},$$

which is also equal to the probability that "A does not win".

Let X be the random variable denoting the amount received by A. Then the spectrum of X is $\{-50, 50\}$ and $P(X=-50)=\frac{5}{11}$, $P(X=50)=\frac{6}{11}$.

Then the expectation of A is given by

$$E(X)=\frac{5}{11}(-50)+\frac{6}{11}(50)=\frac{50}{11} \text{ (in rupees).}]$$

27. Find the mean and the variance of the discrete random variable X with probability mass function given by

$$f(x)=kq^x,\ x=0, 1, 2, \ldots\ (0<q<1)$$

[*Hint* : $\sum_{x=0}^{\infty} kq^x=1$ gives $k=1-q$.

$$m=E(X)=\sum_{x=1}^{\infty} x(1-q)q^x=(1-q)(q+2q^2+3q^3+4q^4+\cdots)$$

$$=q(1-q)(1-q)^{-2}=\frac{q}{1-q}.$$

$$E\{(X-1)X\}=\sum_{x=2}^{\infty} x(x-1)(1-q)q^x$$

$$=(1-q)(2q^2+3\cdot 2q^3+4\cdot 3q^4+\cdots)$$

$$=2q^2(1-q)(1+3q+6q^2+\cdots)$$

$$=2q^2(1-q)(1-q)^{-3}=\frac{2q^2}{(1-q)^2}.$$

$$\text{Var}(X)=E\{X(X-1)\}+E(X)-\{E(X)\}^2$$

$$=\frac{q}{(1-q)^2}.$$

See also 7·4 (Geometric distribution).]

Ex. VII

28. Find the moment generating function of a uniform random variable in $(-a, a)$ and hence find $E(X^n)$. [*C. H.* (*Math.*) '93]

[*Hint* : See Ex. 32 worked out.]

29. For any real characteristic function $\phi(t)$, prove that

$$1+\phi(2t) \geqslant 2\{\phi(t)\}^2.$$

30. Find the probability density function $f(x)$ corresponding to the characteristic function $\phi(t)$ given by

$$\begin{aligned}\phi(t) &= 1-|t| \quad \text{if } |t| \leqslant 1\\ &= 0 \quad \text{if } |t| > 1.\end{aligned}$$

31. Find the median of the binomial $(6, \frac{1}{3})$ distribution.

[*Hint* : See worked out Ex. 25.]

32. A random variate X is uniformly distributed in the interval $(-2, 2)$. If a random variable Z is defined by the formula

$$Z=\begin{cases} X, & \text{if } X<1\\ 1, & \text{if } X \geqslant 1,\end{cases}$$

then find the distribution of the random variable Z and the mean and variance of Z. [*C. H.* (*Math.*) '90]

[*Hint* : See worked out Ex. 40 (ii) and example 3 of section 7·2.]

33. Find the moment generating function of the normal (0, 1) distribution and hence deduce the nth central moments.

[See the concluding part of 7·5.] [*C. H.* (*Math.*) '87]

34. Compute the moment generating function of the distribution defined by the density function

$$f(x)=xe^{-x}, \; x>0.$$ [*C. H.* (*Math.*) '87]

[*Hint* : See 7.6.17.]

35. A motorist encounters n consecutive traffic lights, each likely to be red with probability p or green with probability $q=1-p$. Let $x-1$ be the number of green lights passed by the motorist before being stopped first by a red light. Show that the probability distribution of x is given by pq^{x-1}, $x=1, 2, \ldots, n$ and find the mean and variance of x as $n \to \infty$. [*C. H.* (*Math.*) '80]

36. A fair coin is tossed four times. Let X denote the number of times a head is followed immediately by a tail. Find the variance of X.

[*Hint* : Here the distribution of X is given by
$P(X=0)=\frac{5}{16}$, $P(X=1)=\frac{10}{16}$, $P(X=2)=\frac{1}{16}$.
$\text{Var}(X)=E(X^2)-\{E(X)\}^2=\frac{5}{16}$.]

37. A discrete random variable with probability mass function given by,

$$P(X=0)=\tfrac{1}{2},\ P\{X=(2k-1)\pi\}=\frac{2}{(2k-1)^2\pi^2},\ k=0,\ \pm 1,\ \pm 2,\ldots$$

Show that the characteristic function $\phi(t)$ of X is given by

$$\phi(t)=\tfrac{1}{2}+\frac{4}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos(2k-1)\pi t}{(2k-1)^2}.$$

[*Hint* : $\phi(t)=E(e^{itX})$

$$=e^{it0}\cdot\tfrac{1}{2}+e^{-it\pi}\cdot\frac{2}{\pi^2}+e^{it\pi}\cdot\frac{2}{\pi^2}+e^{-3it\pi}\cdot\frac{2}{3^2\pi^2}$$
$$+e^{3it\pi}\cdot\frac{2}{3^2\pi^2}+e^{-5it\pi}\cdot\frac{2}{5^2\pi^2}+e^{5it\pi}\cdot\frac{2}{5^2\pi^2}+\cdots$$
$$=\tfrac{1}{2}+\frac{2}{\pi^2}(e^{it\pi}+e^{-it\pi})+\frac{2}{3^2\pi^2}(e^{3it\pi}+e^{-3it\pi})$$
$$+\frac{2}{5^2\pi^2}(e^{5it\pi}+e^{-5it\pi})+\cdots$$
$$=\tfrac{1}{2}+\frac{4}{\pi^2}\cos t\pi+\frac{4\cos 3t\pi}{3^2\pi^2}+\frac{4\cos 5t\pi}{5^2\pi^2}$$
$$=\tfrac{1}{2}+\frac{4}{\pi^2}\sum_{k=1}^{\infty}\frac{\cos(2k-1)\pi t}{(2k-1)^2}.]$$

38. The probability density function of a continuous random variable X is given by

$$f(x)=\theta x+\tfrac{1}{2}\ \text{ if } -1<x<1$$
$$=0,\quad \text{elsewhere},$$

where θ is a constant. Find the value (or values) of θ for which $\text{Var}(X)$ is maximum.

Ex. 37

[Hint: $E(X)=\frac{2}{3}\theta$, $E(X^2)=\frac{1}{3}$,

$\text{Var}(X)=\frac{1}{3}-\frac{4}{9}\theta^2=\frac{4}{9}(\frac{3}{4}-\theta^2)$.

Now, in order that $f(x)>0$ in $-1<x<1$, we must have $-\frac{1}{2}<\theta<\frac{1}{2}$. Again, $\frac{3}{4}-\theta^2$ is a maximum when $\theta=0$ which belongs to $(-\frac{1}{2},\frac{1}{2})$. Hence the required value of θ is 0.]

39. Find the mean and the variance of the distribution in Ex. 27 by finding the probability generating function of the corresponding distribution.

[Hint: Here $P(s)=\sum_{x=0}^{\infty} kq^x s^x$, where $k=1-q$,

Then $P(s)=(1-q)\sum_{x=0}^{\infty}(qs)^x$

$$=(1-q)\,\frac{1}{1-qs} \text{ if } |s|<\frac{1}{q}.$$

Now $P'(s)=\dfrac{(1-q)q}{(1-qs)^2}$.

$$\therefore\ P'(1)=\frac{(1-q)q}{(1-q)^2}\quad\left(\frac{1}{q}>|1|=1\right).$$

$$\therefore\ m=E(X)=P'(1)=\frac{q}{1-q}.$$

Also $E\{X(X-1)\}=P''(1)=\dfrac{2q^2(1-q)}{(1-q)^3}=\dfrac{2q^2}{(1-q)^2}$.

So $\text{Var}(X)=\dfrac{2q^2}{(1-q)^2}-\dfrac{q}{(1-q)}\left(\dfrac{q}{1-q}-1\right)$

$$=\frac{q}{(1-q)^2}.\ ]$$

Answers

1. (i) $\frac{4}{3}, \frac{8}{9}$; (ii) 0, 2

5. $\dfrac{2n+1}{3}$.

6. $p, p(1-p), \dfrac{1-2p}{\sqrt{p(1-p)}}, \dfrac{1-3p(1-p)}{\sqrt{p(1-p)}}$.

11. $b, b, \dfrac{b \log_e 3}{2}$.

12. $\frac{1}{4}$.

13. $1, \frac{2}{3}$.

14. $\dfrac{2}{3}, \dfrac{1}{\sqrt{2}}$.

15. $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$.

19. Unique mode $\dfrac{n}{2}$ if n is even and if n be odd, then th[e] modes are $\dfrac{n+1}{2}, \dfrac{n-1}{2}$.

20. mode is $l-1$, $\gamma_1 = \dfrac{2}{\sqrt{l}}$.

21. $\beta_1 = \dfrac{(1-2p)^2}{npq}$, $\beta_2 = \dfrac{1-6pq}{npq} + 3$.

22. $\dfrac{(\frac{1}{2}+l)\Gamma(\frac{1}{2}+l)}{\Gamma(l)}$.

24. Mode does not exist. Median is $\dfrac{1}{\lambda} \log_e 2$.

28. 0 if n is odd and $\dfrac{a^n}{n+1}$ if n is even.

30. $f(x) = \dfrac{1}{\pi} \cdot \dfrac{1-\cos x}{x^2}, \; -\infty < x < \infty$.

31. 2 is the median.

34. $\dfrac{1}{1-t}$ if $t < 1$.

35. $\dfrac{1}{p}, \dfrac{q}{(1-q)^2}$.

11. $b, b, \frac{b \log_e 3}{2}$.

12. $\frac{1}{4}$.

13. $1, \frac{8}{3}$.

14. $\frac{2}{3}, \frac{1}{\sqrt{2}}$.

15. $\frac{1}{3}, \frac{1}{2}, \frac{1}{2}$.

19. Unique mode $\frac{n}{2}$ if n is even and if n be odd, then the modes are $\frac{n+1}{2}, \frac{n-1}{2}$.

20. mode is $l-1$, $\gamma_1 = \frac{2}{\sqrt{l}}$.

21. $\beta_1 = \frac{(1-2p)^2}{npq}$, $\beta_2 = \frac{1-6pq}{npq} + 3$.

22. $\frac{(\frac{1}{2}+l)\Gamma(\frac{1}{2}+l)}{\Gamma(l)}$.

24. Mode does not exist. Median is $\frac{1}{\lambda} \log_e 2$.

28. 0 if n is odd and $\frac{a^n}{n+1}$ if n is even.

30. $f(x) = \frac{1}{\pi} \cdot \frac{1-\cos x}{x^2}$, $-\infty < x < \infty$.

31. 2 is the median.

34. $\frac{1}{1-t}$ if $t < 1$.

35. $\frac{1}{p}, \frac{q}{(1-q)^2}$.

CHAPTER VIII

MATHEMATICAL EXPECTATION—II

8.1. Two Dimensional Expectation.

We consider a two dimensional random variable (X, Y). Let $g(X, Y)$ be a random variable where $g: R \times R \to R$ is a continuous function. Then the expectation of $g(X, Y)$, denoted by $E\{g(X, Y)\}$, is defined in the following cases :

Case I. The distribution of (X, Y) is discrete.

Let $P(X=x_i, Y=y_j)=f_{ij}$, where (x_i, y_j) is a point of the spectrum of (X, Y). If the double series $\sum_i \sum_j g(x_i, y_j) f_{ij}$ [summation is taken over all points of the spectrum of (X, Y)] be absolutely convergent then we say that $E\{g(X, Y)\}$ exists and the value of $E\{g(X, Y)\}$ is given by

$$E\{g(X, Y)\} = \sum_i \sum_j g(x_i, y_j) f_{ij}. \tag{8.1.1}$$

Case II. The distribution of (X, Y) is continuous.

Let $f(x, y)$ be the probability density function of (X, Y). If the double integral $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy$ be absolutely convergent, then we say that $E\{g(X, Y)\}$ exists and the value of $E\{g(X, Y)\}$ is given by

$$E\{g(X, Y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy. \tag{8.1.2}$$

Remark. If $g(X, Y)=g_1(X)$ [or $g_2(Y)$], then the expectation of $g(X, Y)$ [if $E\{g(X, Y)\}$ exists] can also be determined by considering the distribution of X alone (or Y alone) and so it should be proved that the values of $E\{g_1(X)\}$ [or of $E\{g_2(Y)\}$] calculated in two ways are equal. Let us prove that the values of $E\{g_1(X)\}$ calculated in two ways are same when the distribution of (X, Y) is continuous. Let $f(x, y)$ be the joint probability density function of the random variables X and Y.

$$\text{Then} \quad E\{g_1(X)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x) f(x, y)\, dx\, dy.$$

$$= \int_{-\infty}^{\infty} \left\{ g_1(x) \int_{-\infty}^{\infty} f(x, y)\, dy \right\} dx$$

$$= \int_{-\infty}^{\infty} g_1(x) f_X(x)\, dx,$$

where $f_X(x)$ is the marginal probability density function of X.

But $\int_{-\infty}^{\infty} g_1(x) f_X(x)\,dx$ is the value of $E\{g_1(X)\}$ with respect to the distribution of X alone. So the desired result is proved when the distribution of (X, Y) is continuous. We can similarly prove the result when the distribution of (X, Y) is discrete.

Then the values of $E(X)$, $E(Y)$, $E(X^2)$, $E(Y^2)$, $E(X^3)$, $E(Y^3)$, etc. (if they exist) will be uniquely determined whether they are calculated with respect to the joint distribution of X and Y or with respect to the individual distribution of X or Y.

The functions $g_1, g_2, \cdots\cdots, g_n, g$ used in the following theorems are all continuous in $R \times R$.

THEOREM 8.1.1. *(a) If $E\{g_1(X, Y)\}$, $E\{g_2(X, Y)\}, \ldots\ldots, E\{g_n(X, Y)\}$ exist, then $E\{g_1(X, Y) + g_2(X, Y) + \cdots\cdots + g_n(X, Y)\}$*

$$= E\{g_1(X, Y)\} + E\{g_2(X, Y)\} + \cdots\cdots + E\{g_n(X, Y)\}. \tag{8.1.3}$$

(b) If k be a real constant, then $E\{k\,g(X, Y)\} = k\,E\{g(X, Y)\}$, provided $E\{g(X, Y)\}$ exists.

(c) If X, Y are independent variates, then

$$E\{g_1(X)\,g_2(Y)\} = E\{g_1(X)\}\,E\{g_2(Y)\}, \tag{8.1.4}$$

provided $E\{g_1(X)\}$ and $E\{g_2(Y)\}$ exist.

(d) If $g(x, y) \geq 0$ for all $(x, y) \in R \times R$ and $E\{g(X, Y)\}$ exists, then

$$E\{g(X, Y)\} \geq 0.$$

(e) If $g(x, y) \geq 0$ for all $(x, y) \in R \times R$ and $E\{g(X, Y)\} = 0$, then the spectrum of $g(X, Y)$ is $\{0\}$

Proof: (a) Case I. Distribution of (X, Y) is discrete.

Let $P(X = x_i, Y = y_j) = f_{ij}$ where (x_i, y_j) is a point of the spectrum of (X, Y). Then

$E\{g_1(X, Y) + g_2(X, Y) + \cdots\cdots g_n(X, Y)\}$ will exist if

$\sum_i \sum_j \{g_1(x_i, y_j) + g_2(x_i, y_j) + \cdots\cdots + g_n(x_i, y_j)\} f_{ij}$ is absolutely convergent. Here $E\{g_1(X, Y)\}$, $E\{g_2(X, Y)\}, \ldots\ldots, E\{g_n(X, Y)\}$ exist. So $\sum_i \sum_j g_1(x_i, y_j) f_{ij}$, $\sum_i \sum_j g_2(x_i, y_j) f_{ij}, \ldots\ldots, \sum_i \sum_j g_n(x_i, y_j) f_{ij}$ are absolutely convergent and consequently

$\sum_i \sum_j \{g_1(x_i, y_j) + g_2(x_i, y_j) + \cdots\cdots + g_n(x_i, y_j)\} f_{ij}$ is absolutely convergent.

Hence, $E\{g_1(X, Y) + g_2(X, Y) + \cdots\cdots + g_n(X, Y)\}$ exists and

$$E\{g_1(X, Y) + g_2(X, Y) + \cdots\cdots + g_n(X, Y)\}$$
$$= \sum_i \sum_j g_1(x_i, y_j) f_{ij} + \sum_i \sum_j g_2(x_i, y_j) f_{ij} + \cdots\cdots + \sum_i \sum_j g_n(x_i, y_j) f_{ij}$$
$$= E\{g_1(X, Y)\} + E\{g_2(X, Y)\} + \cdots\cdots + E\{g_n(X, Y)\}.$$

Case II. Distribution of (X, Y) is continuous. Let $f(x, y)$ be the joint probability density function of X and Y. Then $E\{g_1(X, Y) + g_2(X, Y) + \cdots\cdots + g_n(X, Y)\}$ will exist if

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{g_1(x, y) + g_2(x, y) + \cdots\cdots\cdots + g_n(x, y)\} f(x, y)\,dx\,dy$ is absolutely convergent. Here $E\{g_1(X, Y)\}$, $E\{g_2(X, Y)\}, \ldots, E\{g_n(X, Y)\}$ exist. So $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x, y) f(x, y)\,dx\,dy$, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(x, y) f(x, y)\,dx\,dy$, $\ldots$, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_n(x, y)\, f(x, y)\,dx\,dy$ are all absolutely convergent and consequently $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{g_1(x, y) + g_2(x, y) + \cdots\cdots + g_n(x, y)\} f(x, y)\,dx\,dy$ is absolutely convergent.

Hence, $E\{g_1(X, Y) + g_2(X, Y) + \cdots\cdots + g_n(X, Y)\}$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{g_1(x, y) + g_2(x, y) + \cdots\cdots + g_n(x, y)\} f(x, y)\,dx\,dy$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x, y) f(x, y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_2(x, y) f(x, y)\,dx\,dy + \cdots$$
$$\cdots\cdots + \int_{-\infty}^{\infty} g_n(x, y) f(x, y)\,dx\,dy$$
$$= E\{g_1(X, Y)\} + E\{g_2(X, Y)\} + \cdots\cdots + E\{g_n(X, Y)\}.$$

(b) Case I. Distribution of (X, Y) is discrete.

Let $P(X = x_i, Y = y_j) = f_{ij}$ where (x_i, y_j) is a point of the spectrum of (X, Y). Then since $E\{g(X, Y)\}$ exists, $\sum_i \sum_j g(x_i, y_j) f_{ij}$ is absolutely convergent and this again implies that

$\sum_i \sum_j k g(x_i, y_j)\, f_{ij} = k \sum_i \sum_j g(x_i, y_j) f_{ij}$ is absolutely convergent, for any real constant k. From this it immediately follows that

$$E\{kg(X, Y)\} = k\,E\{g(X, Y)\}.$$

Case II. Distribution of (X, Y) is continuous. The relation can be proved similarly in this case.

(c) Case I. Distribution of (X, Y) is discrete.

$$E\{g_1(X)\,g_2(Y)\} = \sum_i \sum_j g_1(x_i)\,g_2(y_j)\,f_{ij},$$

where $P(X = x_i, Y = y_j) = f_{ij}$ and (x_i, y_j) is a point of the spectrum of (X, Y), provided the double series is absolutely convergent.

Here $E\{g_1(X)\}$ and $E\{g_2(Y)\}$ exist.

So $\sum_j \sum_i g_1(x_i)\ f_{ij}$ and $\sum_i \sum_j g_2(y_j) f_{ij}$ are absolutely convergent.

So $\sum_i \sum_j g_1(x_i) f_{ij} = \sum_i \{g_1(x_i) \sum_j f_{ij}\}$

$$= \sum_i g_1(x_i) f_{xi}, \text{ since } f_{xi} = \sum_j f_{ij},$$

and $\sum_i \sum_j g_2(y_j) f_{ij} = \sum_j \{g_2(y_j) \sum_i f_{ij}\}$

$$= \sum_j g_2(y_j)\, f_{yj}, \text{ since } f_{yj} = \sum_i f_{ij}.$$

Therefore $\sum_i \sum_j g_1(x_i)\,g_2(y_j)\,f_{xi}\,f_{yj}$ is absolutely convergent. Also here X, Y are independent. So $f_{ij} = f_{xi}\,f_{yj}$, for all i, j and hence $\sum_i \sum_j g_1(x_i)\,g_2(y_j) f_{ij}$ is absolutely convergent. Consequently

$E\{g_1(X)\,g_2(Y)\}$ exists and

$$E\{g_1(X) g_2(Y)\} = \sum_i \sum_j g_1(x_i)\,g_2(y_j)\,f_{xi}\,f_{yj}$$

$$= \{\sum_i g_1(x_i) f_{xi}\} \times \{\sum_j g_2(y_j) f_{yj}\}$$

$$= E\{g_1(X)\}\ E\{g_2(Y)\}.$$

Case II. Distribution of (X, Y) is continuous.

$$E\{g_1(X)\ g_2(Y)\}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)\,g_2(y)\,f(x, y)\ dx\ dy,$$

where $f(x, y)$ is the joint probability density function of X and Y, provided the double integral on the right hand side is absolutely convergent.

Here $E\{g_1(X)\}$ and $E\{g_2(Y)\}$ exist. So

$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)\, f(x, y)\ dx\ dy$ and $\int_{-\infty}^{\infty} g_2(y)\ f(x, y)\ dx\ dy$ are absolutely convergent.

Now $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)\ f(x, y)\ dx\ dy$

$$= \int_{-\infty}^{\infty} g_1(x) \left\{\int_{-\infty}^{\infty} f(x, y)\ dy\right\} dx$$

$$= \int_{-\infty}^{\infty} g_1(x) f_X(x)\ dx$$

and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_2(y)\, f(x, y)\ dx\ dy$

$$= \int_{-\infty}^{\infty} g_2(y) \left\{\int_{-\infty}^{\infty} f(x, y)\ dx\right\} dy$$

$$= \int_{-\infty}^{\infty} g_2(y)\ f_Y(y)\, dy.$$

$\therefore \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)\,g_2(y)\,f_X(x)\,f_Y(y)\,dx\ dy$ is absolutely convergent.

Also here X, Y are independent. So $f(x, y) = f_X(x)\, f_Y(y)$, for all x, y.

So $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x)\,g_2(y)\ f(x, y)\ dx\ dy$ is absolutely convergent.

and $E\{g_1(X)\ g_2(Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x) g_2(y)\, f_X(x)\, f_Y(y)\ dx\ dy$

$$= \left(\int_{-\infty}^{\infty} g_1(x)\, f_X(x)\ dx\right)\left(\int_{-\infty}^{\infty} g_2(y)\, f_Y(y)\ dy\right)$$

$$= E\{g_1(X)\}\ E\{g_2(Y)\}.$$

(d) Case I. Distribution of (X, Y) is discrete.

Here $\sum_i \sum_j g(x_i, y_j)\, f_{ij}$ is absolutely convergent.

Also here $g(x_i, y_j) \geq 0$, $f_{ij} > 0$ for all i, j.

So $\sum_i \sum_j g(x_i, y_j)\, f_{ij} \geq 0$ and hence,

$E\{g(X, Y)\} \geq 0$.

Case II. Distribution of (X, Y) is continuous.

Here $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f(x, y)$ is absolutely convergent.

Also here $g(x, y) \geq 0$, $f(x, y) \geq 0$ for all x, y.

So $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f(x, y)\ dx\ dy \geq 0$ and hence

$E\{g(X, Y)\} \geq 0$.

(e) Case I. Distribution of (X, Y) is discrete.

Here $E\{g(X, Y)\}$ exists.

Also $E\{g(X, Y)\} = \sum_i \sum_j g(x_i, y_j) f_{ij} = 0$.

But $f_{ij} > 0$ and $g(x_i, y_j) \geqslant 0$ for all i, j.

Then $\sum_i \sum_j g(x_i, y_j) f_{ij} = 0$ implies that $g(x_i, y_j) = 0$ for every point (x_i, y_j) of the spectrum of (X, Y) and consequently the spectrum of $g(X, Y)$ is $\{0\}$.

Case II. Distribution of (X, Y) is continuous.

Let $f(x, y)$ be the probability density function of (X, Y).

Hence $E\{g(X, Y)\}$ exists and

$$E\{g(X, Y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy = 0.$$

But $f(x, y) \geqslant 0$ and $g(x, y) \geqslant 0$ for all x, y and g is continuous in $R \times R$.

So $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy = 0$ implies that $g(x, y) = 0$ for every point (x, y) of the spectrum of (X, Y) and consequently the spectrum of $g(X, Y)$ is $\{0\}$.

Cor. 1. $E(X+Y) = E(X) + E(Y)$ if $E(X)$ and $E(Y)$ exist. (8.1.5)

Taking $n = 2$ and $g_1(X, Y) = X$, $g_2(X, Y) = Y$ we find from (a),

$$\begin{aligned} E\{g_1(X, Y) + g_2(X, Y)\} &= E(X+Y) \\ &= E\{g_1(X, Y)\} + E\{g_2(X, Y)\} \\ &= E(X) + E(Y). \end{aligned}$$

Cor. 2. If X, Y are independent random variables, then taking $g_1(X) = X$, $g_2(Y) = Y$, we find from (c) that $E(XY) = E(X)\,E(Y)$, provided $E(Y)$, $E(X)$ exist.

8.2. Moments, Covariance and Correlation Coefficient.

Moments. Let X, Y be two random variables and let r, k be any two given non-negative integers. If $E(X^r Y^k)$ exists, then it is called a bivariate moment of order $(r+k)$ about the origin and is denoted by α_{rk}. So we have $\alpha_{00} = 1$, $\alpha_{11} = E(XY)$, $\alpha_{10} = E(X)$, $\alpha_{01} = E(Y)$, $\alpha_{32} = E(X^3 Y^2)$ etc., provided the expectations exist. In general, if $E\{(X-a)^r (Y-b)^k\}$ exists then it will be called a moment of order $(r+k)$ about the point (a, b). We note that α_{10}, α_{01} are respectively

the means of the distributions of X and Y. We denote these means as m_x and m_y respectively. Then $E\{(X-m_x)^r (Y-m_y)^k\}$ is called a $(r+k)$th order bivariate central moment and is denoted by μ_{rk}, provided $E\{(X-m_x)^r (Y-m_y)^k\}$ exists. So we have $\mu_{00} = 1$, $\mu_{10} = \mu_{01} = 0$, $\mu_{11} = E\{(X-m_x)(Y-m_y)\}$, $\mu_{12} = E\{(X-m_x)(Y-m_y)^2\}$, etc., provided the expectations exist. Now if the standard deviations of X and Y exist and if they are denoted by σ_x, σ_y respectively, we find that $\mu_{20} = E\{(X-m_x)^2\} = \sigma_x^2$ and $\mu_{02} = E\{(Y-m_y)^2\} = \sigma_y^2$.

Covariance. Let X, Y be two random variables and let m_x, m_y exist. In many situations it will be essential to investigate the type of probable relationship between the values of the random variables X and Y. It will be explained in section 8.3, that a measure of 'tendency of having linear relationship between X and Y' is given by $E\{(X-m_x)(Y-m_y)\}$, provided the expectation exists. So the quantity μ_{11} gives a measure of the aforesaid property of a bivariate distribution and it is called the covariance between X and Y and is denoted by cov (X, Y). Thus the covariance between X and Y is defined by

$$\text{cov}\,(X, Y) = E\{(X-m_x)(Y-m_y)\} \tag{8.2.1}$$

provided the expectation on the right hand side exists.

Correlation Coefficient.

We observe that for a bivariate distribution, Cov (X, Y) is a measure of a property which does not depend on the units of measurement. But cov (X, Y) is not dimensionless. Keeping this in view, we define the correlation coefficient between X and Y, denoted by $\rho(X, Y)$, and defined by

$$\rho(X, Y) = \frac{\text{Cov}\,(X, Y)}{\sigma_x \sigma_y} \tag{8.2.2}$$

where $\sigma_x (> 0)$, $\sigma_y (> 0)$ are respectively the standard deviations of X and Y.

The following examples illustrate how $E(XY)$, $E(X+Y)$, cov (X, Y) can be computed for a given joint distribution.

Ex. 1. *Two balls are drawn without replacement from an urn containing three balls, numbered 1, 2, 3. Let X be the random variable denoting the number on the first ball drawn and Y the larger of the two numbers drawn. Find* cov (X, Y).

Here the spectrum of X and Y are respectively $\{1, 2, 3\}$ and $\{2, 3\}$. The joint distribution of X and Y is given by

$$P(X=1, Y=2)=\tfrac{1}{3}\cdot\tfrac{1}{2}=\tfrac{1}{6}$$
$$P(X=1, Y=3)=\tfrac{1}{3}\cdot\tfrac{1}{2}=\tfrac{1}{6}$$
$$P(X=2, Y=2)=\tfrac{1}{3}\cdot\tfrac{1}{2}=\tfrac{1}{6}$$
$$P(X=2, Y=3)=\tfrac{1}{3}\cdot\tfrac{1}{2}=\tfrac{1}{6}$$

$P(X=3, Y=2)=0$, since $(X=3, Y=2)$ is an impossible event

$$P(X=3, Y=3)=\tfrac{1}{3}\cdot\tfrac{3}{3}=\tfrac{1}{3}$$

The marginal distributions of X and Y are given by

$$P(X=1)=\tfrac{1}{3},\ P(X=2)=\tfrac{1}{3}\ \ P(X=3)=\tfrac{1}{3}$$
$$P(Y=2)=\tfrac{1}{6}+\tfrac{1}{6}+0=\tfrac{1}{3}\ \ P(Y=3)=\tfrac{1}{3}+\tfrac{1}{6}+\tfrac{1}{6}=\tfrac{2}{3}.$$

Then $m_x=E(X)=1\cdot\tfrac{1}{3}+2\cdot\tfrac{1}{3}+3\cdot\tfrac{1}{3}=2$

$$m_y=E(Y)=2\times\tfrac{1}{3}+3\times\tfrac{2}{3}=\tfrac{8}{3}$$

$$E(XY)=1\cdot2\cdot\tfrac{1}{6}+1\cdot3\cdot\tfrac{1}{6}+2\cdot2\cdot\tfrac{1}{6}+2\cdot3\cdot\tfrac{1}{6}+3\cdot2\cdot0$$
$$+3\cdot3\cdot\tfrac{1}{3}=\tfrac{1}{3}+\tfrac{1}{2}+\tfrac{2}{3}+1+3=\tfrac{11}{2}.$$

$\therefore$ $\operatorname{cov}(X, Y)=E(XY)-m_x m_y$ [see (8.2.5)]

$$=\tfrac{11}{2}-\tfrac{16}{3}=\tfrac{1}{6}.$$

Ex. 2. *Let the joint distribution of X and Y be given by the probability density function*

$$f(x, y)=x+y,\ \text{if } 0<x<1,\ 0<y<1$$
$$=0,\ \text{elsewhere.}$$

Find (*i*) $E(XY)$, (*ii*) $E(X+Y)$.

$$E(XY)=\int_0^1\int_0^1 xy(x+y)\,dx\,dy$$
$$=\int_0^1\left(\frac{x^2}{2}+\frac{x}{3}\right)dx$$
$$=\tfrac{1}{6}+\tfrac{1}{6}$$
$$=\tfrac{1}{3}.$$

$$E(X+Y)=\int_0^1\int_0^1 (x+y)^2\,dx\,dy$$
$$=\int_0^1 (x^2+x+\tfrac{1}{3})\,dx$$
$$=\tfrac{1}{3}+\tfrac{1}{2}+\tfrac{1}{3}$$
$$=\tfrac{7}{6}.$$

Important Properties and Results of Covariance and Correlation Coefficient.

THEOREM 8.2.1. $\operatorname{cov}(X, Y)=\operatorname{cov}(Y, X)$. (8.2.3)

Proof: $\operatorname{cov}(X, Y)=E\{(X-m_x)(Y-m_y)\}=E\{(Y-m_y)(X-m_x)\}$
$=\operatorname{cov}(Y, X)$.

Cor. $\rho(X, Y)=\rho(Y, X)$, if $\sigma_x>0$, $\sigma_y>0$. (8.2.4)

THEOREM 8.2.2. $\operatorname{cov}(X, Y)=E(XY)-m_x m_y$, (8.2.5)

provided $E(XY)$ exists.

Proof: We have $E\{(X-m_x)(Y-m_y)\}$

$$=E(XY-m_xY-m_yY+m_xm_y)$$
$$=E(XY)-m_x E(Y)-m_y E(X)+m_xm_y,$$

since $E(X)$, $E(Y)$ and $E(XY)$ exist,

$$=E(XY)-m_xm_y.$$

Hence, $\operatorname{cov}(X, Y)=E(XY)-m_x m_y$.

THEOREM 8.2.3. *If $a(\neq 0)$, $c(\neq 0)$, b, d are constants, then*

$$\rho(aX+b, cY+d)=\frac{ac}{|a||c|}\rho(X, Y). \tag{8.2.6}$$

Proof: $E(aX+b)=aE(X)+b=am_x+b$

$E(cY+d)=cE(Y)+d=cm_y+d$.

Then $\operatorname{cov}(aX+b, cY+d)$

$$=E\{(aX+b-am_x-b)(cY+d-cm_y-d)\}$$
$$=E\{ac(X-m_x)(Y-m_y)\}$$
$$=ac\,E\{(X-m_x)(Y-m_y)\}$$
$$=ac\operatorname{cov}(X, Y).$$

Again $\operatorname{var}(aX+b)=a^2\ \operatorname{var}(X)=a^2\sigma_x^2$

$\operatorname{var}(cY+d)=c^2\ \operatorname{var}(Y)=c^2\sigma_y^2$.

$\therefore$ standard deviation of $aX+b$ is $|a|\,\sigma_x$ and that of $cY+d$ is $|c|\,\sigma_y$.

So $\rho(aX+b,\, cY+d)$

$$= \frac{\text{cov}\,(aX+b,\, cY+d)}{|a|\sigma_x\,|c|\sigma_y}$$

$$= \frac{ac}{|a|\,|c|}\,\frac{\text{cov}\,(X,\, Y)}{\sigma_x\,\sigma_y}$$

$$= \frac{ac}{|a|\,|c|}\,\rho(X,\, Y).$$

Hence the theorem.

Cor. 1. If $a > 0, c > 0$, we find that

$$\rho(aX+b\,,\, cY+d) = \rho(X,\, Y). \tag{8.2.7}$$

which signifies that the correlation coefficient between two random variables is independent of the units of measurement and the choice of orign.

Cor. 2. $\rho(aX+b,\, cY+d) = \rho(X,\, Y)$ or, $-\rho(X,\, Y)$, according as a, c have the same sign or opposite signs.

THEOREM 8.2.4. *If the correlation coefficient $\rho(X, Y)$ between two random variables X and Y exists, then*

$$-1 \leq \rho(X,\, Y) \leq 1. \tag{8.2.8}$$

Proof: Let $g(x,\, y) = \dfrac{x-m_x}{\sigma_x} + \dfrac{y-m_y}{\sigma_y}$.

Here $g(x, y)$ is a real valued continuous function defined in $R\times R$. Now

$$\left(\frac{x-m_x}{\sigma_x} + \frac{y-m_y}{\sigma_y}\right)^2 \geq 0 \text{ for all } (x,\, y) \in R\times R.$$

Also $E\left\{\left(\dfrac{X-m_x}{\sigma_x} + \dfrac{Y-m_y}{\sigma_y}\right)^2\right\}$ exists, since $\rho(X, Y)$ exists.

Then by (*d*) of Theorem 8.1.1, we get

$$E\left\{\left(\frac{X-m_x}{\sigma_x} + \frac{Y-m_y}{\sigma_y}\right)^2\right\} \geq 0$$

or, $E\left\{\left(\dfrac{X-m_x}{\sigma_x}\right)^2\right\} + E\left\{\left(\dfrac{Y-m_y}{\sigma_y}\right)^2\right\} + 2E\left\{\dfrac{(X-m_x)(Y-m_y)}{\sigma_x\,\sigma_y}\right\} \geq 0$

or, $\dfrac{\sigma_x^2}{\sigma_x^2} + \dfrac{\sigma_y^2}{\sigma_y^2} + 2\rho(X,\, Y) \geq 0$ (8.2.9)

or, $1 + \rho(X,\, Y) \geq 0$.

Taking $g_1(x,\, y) = \dfrac{x-m_x}{\sigma_x} - \dfrac{y-m_y}{\sigma_y}$, $(x, y) \in R\times R$, it can be proved similarly that

$$1 - \rho(X,\, Y) \geq 0. \tag{8.2.10}$$

From (8.2.9) and (8.2.10) we conclude that

$$-1 \leq \rho(X,\, Y) \leq 1.$$

Uncorrelated Random Variables.

Two random variables X, Y are said to be uncorrelated if and only if

$$\text{cov}\,(X,\, Y) = 0. \tag{8.2.11}$$

So if $\sigma_x > 0$, $\sigma_y > 0$, the random variables X, Y are uncorrelated if and only if

$$\rho(X,\, Y) = 0. \tag{8.2.12}$$

THEOREM 8.2.5. *If X, Y are independent random variables, then*

$$\rho(X,\, Y) = 0,$$

provided $\rho(X, Y)$ exists.

Proof: Here cov (X, Y) exists, since $\rho(X, Y)$ exists.

Now, cov (X, Y)

$$= E\{(X-m_x)(Y-m_y)\}$$

$$= E\{(X-m_x)\}\, E\{(Y-m_y)\}, \text{ by } (c) \text{ of Theorem } 8.1.1$$

$$= \{E(X)-m_x\}\{E(Y)-m_y\}$$

$$= 0.$$

Hence $\rho(X,\, Y) = 0$.

Note. The following example shows that the converse of the above theorem is not true.

Let the joint probability distribution of the discrete random variables X, Y be given by

$$P(X=0,\, Y=0) = \tfrac{1}{4},\quad P(X=1,\, Y=0) = \tfrac{1}{4},$$
$$P(X=-1,\, Y=0) = \tfrac{1}{4},\quad P(X=0,\, Y=1) = \tfrac{1}{4}.$$

Here the marginal distributions of X and Y are given by

$$P(X=0) = \tfrac{1}{2},\quad P(X=1) = \tfrac{1}{4},\quad P(X=-1) = \tfrac{1}{4}.$$
$$P(Y=0) = \tfrac{3}{4},\quad P(Y=1) = \tfrac{1}{4}.$$

Then $m_x = 0\,.\,\tfrac{1}{2} + 1\,.\,\tfrac{1}{4} + (-1)\,\tfrac{1}{4} = 0$

$m_y = 0\,.\,\tfrac{3}{4} + 1\,.\,\tfrac{1}{4} = \tfrac{1}{4}.$

Also $E(XY) = 0 \cdot 0 \cdot \frac{1}{4} + 1 \cdot 0 \cdot \frac{1}{4} + (-1) \cdot 0 \cdot \frac{1}{4} + 0 \cdot 1 \cdot \frac{1}{4} = 0.$

So $\operatorname{cov}(X, Y) = E(XY) - m_x m_y = 0.$

Hence, $\rho(X, Y) = 0.$

Now we see that

$$P(X=0,\ Y=0) = \tfrac{1}{4},\ P(X=0) = \tfrac{1}{2},\ P(Y=0) = \tfrac{3}{4}.$$

Therefore, $P(X=0,\ Y=0) \neq P(X=0)\, P(Y=0)$, and consequently X, Y are not independent. So for the above random variables X, Y we find that $\rho(X, Y) = 0$ but X, Y are not independent.

THEOREM 8.2.6. **Cauchy-Schwarz Inequality.**

If X and Y be two random variables such that $E(X^2)$, $E(Y^2)$ and $E(XY)$ exist, then $\{E(XY)\}^2 \leq E(X^2)\, E(Y^2)$, (8.2.13)

where the equality sign holds if and only if $E(X^2) = 0$ or $P(Y - aX = 0) = 1$ for some constant a.

Proof: Let $g(x, y) = (y - kx)^2$ where k is a real constant and $(x, y) \in R \times R$. Then $g(x, y)$ is a real valued continuous function defined in $R \times R$. Now

$$E\{g(X, Y)\} = E\{(Y - kX)^2\},$$

which exists, since $E(X^2)$, $E(Y^2)$ and $E(XY)$ exist.

Also, $g(x, y) = (y - kx)^2 \geq 0$ for all $(x, y) \in R \times R$.

Hence, $E\{(Y - kX)^2\} \geq 0$

Now, $E\{(Y - kX)^2\}$

$$= E(Y^2 - 2kXY + k^2 X^2)$$

$$= E(Y^2) - 2k\, E(XY) + k^2 E(X^2).$$

So we get $E(Y^2) - 2k\, E(XY) + k^2 E(X^2) \geq 0$. (8.2.14)

for all real values of k.

Now, there are two possibilities :

(*i*) $E(X^2) = 0$, (*ii*) $E(X^2) \neq 0$.

In case (*i*), by (*e*) of Theorem 8.1.1, the spectrum of X is the set $\{0\}$, where $P(X=0) = 1$. So in this case $E(X) = 0$ and $E(XY) = 0$ and consequently the required inequality is established with the sign of equality.

In case (*ii*), $E(X^2) > 0$ and so (8.2.14) gives

$$k^2 - 2k\frac{E(XY)}{E(X^2)} + \frac{E(Y^2)}{E(X^2)} \geq 0 \text{ for all real } k.$$

$$\text{or, } \left\{k - \frac{E(XY)}{E(X^2)}\right\}^2 + \frac{E(Y^2)}{E(X^2)} - \left\{\frac{E(XY)}{E(X^2)}\right\}^2 \geq 0 \qquad (8.2.15)$$

for all real k.

Now, $\dfrac{E(XY)}{E(X^2)}$ is a real number.

So, taking $k = \dfrac{E(XY)}{E(X^2)}$ in (8.2.15), we get

$$\frac{E(Y^2)}{E(X^2)} - \left\{\frac{E(XY)}{E(X^2)}\right\}^2 \geq 0$$

or, $\dfrac{E(Y^2)}{E(X^2)} \geq \left\{\dfrac{E(XY)}{E(X^2)}\right\}^2$

or, $E(X^2)\, E(Y^2) \geq \{E(XY)\}^2$, since $E(X^2) > 0$.

Hence, $[E(XY)]^2 \leq E(X^2)\, E(Y^2)$

In case (*ii*) we find that

$$\{E(XY)\}^2 = E(X^2)\, E(Y^2)$$

if and only if the roots of $k^2 E(X^2) - 2k\, E(XY) + E(Y^2) = 0$ are real and equal, *i.e.*, if and only if there exists a real value of k, say a, for which $E\{(Y - aX)^2\} = 0$. Now $E\{(Y - aX)^2\} = 0$ if and only if the spectrum of the random variable $Y - aX$ is the set $\{0\}$ with $P(Y - aX = 0) = 1$. Hence, it is proved that in case (*ii*),

$$\{E(XY)\}^2 = E(X^2)\, E(Y^2)$$

if and only if $P(Y - aX = 0) = 1$ for some real constant a. Also it has been proved that in case (*i*), $\{E(XY)\}^2 = E(X^2)\, E(Y^2)$, when the value of each side is 0.

Thus in any case the required inequality is established and further $\{E(XY)\}^2 = E(X^2)\, E(Y^2)$ if and only if $E(X^2) = 0$ or $P(Y - aX = 0) = 1$ for some real constant a.

Cor. $-1 \leq \rho(X, Y) \leq 1$,

which is already established in theorem 8.2.4.

We consider the random variables X^*, Y^* defined by

$$X^* = \frac{X - m_x}{\sigma_x},\quad Y^* = \frac{Y - m_y}{\sigma_y}.$$

Now applying Cauchy-Schwarz inequality to X^*, Y^*, we get

$$\{E(X^*Y^*)\}^2 \le E\{(X^*)^2\} E\{(Y^*)^2\}. \tag{8.2.16}$$

But $E\{(X^*)^2\} = E\left\{\left(\frac{X-m_x}{\sigma_x}\right)^2\right\} = 1$

and $E\{(Y^*)^2\} = E\left\{\left(\frac{Y-m_y}{\sigma_y}\right)^2\right\} = 1.$

Also, $E(X^*Y^*) = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} = \rho(X, Y).$

So, from (8.2.16) we find that

$$\{\rho(X, Y)\}^2 \le 1.$$

Hence, $-1 \le P(X, Y) \le 1.$

THEOREM 8.2.7. *If X, Y are two random variables having standard deviations σ_x, σ_y respectively and if cov (X, Y) exists, then*

$$\text{var}(aX+bY) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab \text{ cov}(X, Y), \tag{8.2.17}$$

where a, b are real constants.

Proof: $E(aX+bY) = aE(X) + bE(Y) = am_x + bm_y.$

Then $\text{var}(aX+bY) = E[\{(aX+bY) - (am_x + bm_y)\}^2]$

$$= E[\{a(X-m_x) + b(Y-m_y)\}^2],$$

which exists since var (X), var (Y) and cov (X, Y) exist.

Now $E[\{a(X-m_x) + b(Y-m_y)\}^2]$

$$= a^2 E\{(X-m_x)^2\} + b^2 E\{(Y-m_y)^2\} + 2ab\, E\{(X-m_x)(Y-m_y)\}$$

$$= a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab \text{ cov}(X, Y).$$

Hence, it is proved that

$$\text{Var}(aX+bY) = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab \text{ cov}(X, Y).$$

Cor. 1. (8.2.17) can be expressed as

$$\text{var}(aX+bY) = a^2 \text{ var}(X) + b^2 \text{ var}(Y) + 2ab\,\rho\sigma_x\sigma_y \tag{8.2.18}$$

if $\sigma_x > 0$, $\sigma_y > 0$.

Cor. 2. Taking $a = b = 1$ and $a = 1$, $b = -1$, we get from (8.2.17)

$$\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \pm 2 \text{ cov}(X, Y). \tag{8.2.19}$$

Cor. 3. If X, Y are uncorrelated, then

$$\text{var}(aX+bY) = a^2 \text{ var}(X) + b^2 \text{ var}(Y). \tag{8.2.20}$$

8.3. Moment Generating Function.

The joint moment generating function $M_{X,Y}(t_1, t_2)$ of the random variables X, Y is defined by

$$M_{X,Y}(t_1, t_2) = E(e^{t_1X + t_2Y}) \tag{8.3.1}$$

provided the expectation exists and where t_1, t_2 are real variables.

We assume that (8.3.1) can be differentiated under expectation partially with respect to t_1 and t_2 any number of times and so if the $(r+k)$th order moment $E(X^rY^k)$ exists, then this moment can be obtained by taking the value of

$$\frac{\partial^{r+k} M(t_1, t_2)}{\partial t_1^{\,r}\, \partial t_2^{\,k}} \text{ at } t_1 = 0,\ t_2 = 0, \text{ where } M = M(t_1, t_2)$$

stands for $M_{X,Y}(t_1, t_2)$. Thus

$$\left[\frac{\partial^{r+k} M}{\partial t_1^{\,r} \partial t_2^{\,k}}\right]_{(t_1=0),(t_2=0)} = E(X^rY^k), \tag{8.3.2}$$

$$\left[\frac{\partial^2 M}{\partial t_1^2}\right]_{(t_1=0,\ t_2=0)} = E(X^2), \tag{8.3.3}$$

$$\left[\frac{\partial^2 M}{\partial t_2^2}\right]_{(t_1=0,\ t_2=0)} = E(Y^2), \tag{8.3.4}$$

$$\left[\frac{\partial^2 M}{\partial t_1 \partial t_2}\right]_{(t_1=0,\ t_2=0)} = E(XY), \tag{8.3.5}$$

etc.

We state (without proof) below an important theorem.

THEOREM 8.3.1. *The joint moment generating function of a bivariate normal distribution with parameters $m_x, m_y, \sigma_x, \sigma_y, \rho$, is given by* $M_{X,Y}(t_1, t_2) = e^{[t_1m_x + t_2m_y + \frac{1}{2}(t_1^2\sigma_x^2 + \rho t_1t_2\sigma_x\sigma_y + t_2^2\sigma_y^2)]}$ (8.3.6)

Now we shall find $E(X)$, $E(Y)$, var (X), var (Y), cov (X, Y), $\rho(X, Y)$ with the help of (8.3.6), where (X, Y) has a bivariate normal distribution with probability density function given by

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}\,\sigma_x\,\sigma_y} \times e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right) + \left(\frac{y-m_y}{\sigma_y}\right)^2\right\}},$$

$$-\infty < x < \infty,\ -\infty < y < \infty,$$

and $\sigma_x > 0$, $\sigma_y > 0$, $-1 < \rho < 1$.

By (8.3.6).

$$\frac{\partial M}{\partial t_1}=(m_x+t_1\sigma_x^2+\rho t_2\sigma_x\sigma_y)\times e^{\{t_1m_x+t_2m_y+\frac{1}{2}(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\}}$$

$$\therefore\quad\left[\frac{\partial M}{\partial t_1}\right]_{(t_1=0,\,t_2=0)}=m_x.$$

So, $E(X)=m_x$.

Similarly, $E(Y)=m_y$.

Again, $$\frac{\partial^2 M}{\partial t_1^2}=\{\sigma_x^2+(m_x+t_1\sigma_x^2+\rho t_2\sigma_x\sigma_y)^2\}\times e^{\{t_1m_x+t_2m_y+\frac{1}{2}(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\}}$$

$$\therefore\quad E(X^2)=\left(\frac{\partial^2 M}{\partial t_1^2}\right)_{(t_1=0,\,t_2=0)}=\sigma_x^2+m_x^2.$$

Similarly, $E(Y^2)=\sigma_y^2+m_y^2$.

So, $\operatorname{var}(X)=E(X^2)-m_x^2=\sigma_x^2$

and $\operatorname{var}(Y)=E(Y^2)-m_y^2=\sigma_y^2$.

Now, $$E(XY)=\left(\frac{\partial^2 M}{\partial t_1\partial t_2}\right)_{(t_1=0,\,t_2=0)}$$

$$=\left[\{\rho\sigma_x\sigma_y+(m_x+t_1\sigma_x^2+\rho t_2\sigma_x\sigma_y)(m_y+t_2\sigma_y^2+\rho t_1\sigma_x\sigma_y)\}\times e^{\{t_1m_x+t_2m_y+\frac{1}{2}(t_1^2\sigma_x^2+2\rho t_1t_2\sigma_x\sigma_y+t_2^2\sigma_y^2)\}}\right]_{(t_1=0,\,t_2=0)}$$

$$=\rho\sigma_x\sigma_y+m_xm_y.$$

Hence, $\operatorname{cov}(X,Y)=E(XY)-E(X)E(Y)$

$$=\rho\sigma_x\sigma_y+m_xm_y-m_xm_y$$

$$=\rho\sigma_x\sigma_y.$$

So, $$\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_x\sigma_y}=\rho.$$

THEOREM 8.3.2. *If (X, Y) has a bivariate normal distribution, then X, Y are independent if and only if X, Y are uncorrelated.*

Proof: We have seen before that for any bivariate distribution $\operatorname{cov}(X, Y)=0$ if X, Y are independent. So also for a bivariate normal distribution $\operatorname{cov}(X, Y)=0$ and consequently $\rho(X, Y)=0$ if X, Y are independent.

Now let $\rho(X, Y)=0$ for a bivariate normal distribution with parameters $m_x, m_y, \sigma_x, \sigma_y, \rho$. But here $\rho(X, Y)=\rho$, so $\rho=0$. Then the joint probability density function of X and Y is given by

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\frac{1}{2}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2+\left(\frac{y-m_y}{\sigma_y}\right)^2\right\}}$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_x}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}\cdot\frac{1}{\sqrt{2\pi}\,\sigma_y}e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}$$

for $-\infty<x<\infty,\ -\infty<y<\infty$.

Now the marginal probability density functions of X and Y are given respectively by

$$f_X(x)=\frac{1}{\sqrt{2\pi}\,\sigma_x}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}},\quad -\infty<x<\infty$$

and $$f_Y(y)=\frac{1}{\sqrt{2\pi}\,\sigma_y}e^{-\frac{(y-m_y)^2}{2\sigma_y^2}},\quad -\infty<y<\infty.$$

So, $f(x, y)=f_X(x)\,f_Y(y)$, for all x, y.

Hence, X, Y are independent.

So the theorem is proved.

8.4. Extension of the Concept of Expectation with respect to n-Dimensional Distribution.

Expectation :

Let $(X_1, X_2, \ldots\ldots, X_n)$ be an n-dimensional random variable and let $g: R^n \to R$ be a continuous function, where $R^n=R\times R\times\cdots\cdots\times R$ (n copies of R).

Then the expectation of $g(X_1, X_2, \ldots\ldots, X_n)$, denoted by $E\{g(X_1, X_2, \ldots\ldots, X_n)\}$, is defined in the following cases :

Case I. The distribution of $(X_1, X_2, \ldots\ldots, X_n)$ is discrete.

Let $f_{X_1, X_2, \ldots\ldots, X_n}(x_1, x_2, \ldots, x_n)=P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n)$, where $(x_1, x_2, \ldots\ldots, x_n)$ is a point of the spectrum of $(X_1, X_2, \ldots\ldots, X_n)$. If $\Sigma\, g(x_1, x_2, \ldots\ldots, x_n)\, f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots\ldots, x_n)$ [summation is taken over all points of the spectrum of $(X_1, X_2, \ldots\ldots, X_n)$] be absolutely convergent, then we say that $E\{g(X_1, \ldots\ldots, X_n)\}$ exists and it is given by $E\{g(X_1, X_2, \ldots, X_n)\}$

$$=\Sigma\, g(x_1, x_2, \ldots, x_n)\, f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \tag{8.4.1}$$

Case II. The distribution of $(X_1, X_2, \ldots, X_n)$ is continuous. Let $f(x_1, x_2, \ldots, x_n)$ be the probability density function of $(X_1, X_2, \ldots, X_n)$. If the multiple integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n$$

be absolutely convergent, then we say that $E\{g(X_1, X_2, \ldots, X_n)\}$ exists and $E\{g(X_1, X_2, \ldots, X_n)\}$ is given by

$$E\{g(X_1, X_2, \ldots, X_n)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n. \quad (8.4.2)$$

We state (without proof) some important theorems.

THEOREM 8.4.1. *If* $E\{g_1(X_1, X_2, \ldots, X_n)\}$, $E\{g_2(X_1, X_2, \ldots, X_n)\}$, ..., $E\{g_k(X_1, X_2, \ldots, X_n)\}$ *exist, then*

$$E\{g_1(X_1, X_2, \ldots, X_n) + g_2(X_1, X_2, \ldots, X_n) + \cdots + g_k(X_1, X_2, \ldots, X_n)\}$$
$$= E\{g_1(X_1, X_2, \ldots, X_n)\} + E\{g_2(X_1, X_2, \ldots, X_n)\} + \cdots + E\{g_k(X_1, X_2, \ldots, X_n)\} \quad (8.4.3)$$

Theorem 8.4.1 is the extension of Theorem 8.1.1.(a) to n-dimensional distribution.

THEOREM 8.4.2. *If* $E\{g(X_1, X_2, \ldots, X_n)\}$ *exists and c is a real constant, then* $E\{cg(X_1, X_2, \ldots, X_n)\} = c\,E\{g(X_1, X_2, \ldots, X_n)\}$. (8.4.4)

Applying Theorem 8.4.1 and Theorem 8.4.2 to $a_1X_1, a_2X_2, \ldots, a_nX_n$ where $a_1, a_2, \ldots, a_n$ are real constants, we get

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n), \quad (8.4.5)$$

provided $E(X_1), E(X_2), \ldots, E(X_n)$ exist.

THEOREM 8.4.3. *If* $X_1, X_2, \ldots, X_n$ *are mutually independent random variables and* $E\{g_1(X_1)\}, E\{g_2(X_2)\}, \ldots, E\{g_n(X_n)\}$ *exist, then*

(i) $$E\{g_1(X_1)\, g_2(X_2) \ldots g_n(X_n)\} = E\{g_1(X_1)\}\, E\{g_2(X_2)\} \ldots E\{g_n(X_n)\} \quad (8.4.6)$$

(ii) $$E\{g_{i_1}(X_{i_1})\, g_{i_2}(X_{i_2}) \ldots g_{i_m}(X_{i_m})\} = E\{g_{i_1}(X_{i_1})\} E\{g_{i_2}(X_{i_2})\} \ldots E\{g_{i_m}(X_{i_m})\}, \quad (8.4.7)$$

where $i_1, i_2, \ldots, i_m$ *are distinct and* $i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, n\}$ *and* $m < n$.

We note that (8.4.6) is a particular case of (8.4.7) if we take $m = n$ and $i_1 = 1, i_2 = 2, \ldots, i_m = i_n = n$.

THEOREM 8.4.4. *If* $var(X_1), var(X_2), \ldots, var(X_n)$ *exist and further* $cov(X_i, X_j)$ *exists for any pair of random variables* X_i, X_j $(i, j = 1, 2, \ldots, n$ and $i \neq j)$, *then*

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= \sum_{i=1}^{n} a_i^2\, var(X_i) + 2\sum_{i<j}\sum a_i a_j\, cov(X_i, X_j), \quad (8.4.8)$$

where $a_1, a_2, \ldots, a_n$ *are real constants.*

(8.4.8) is an extension of (8.2.17) to n-dimension.

Cor. *If* $X_1, X_2, \ldots, X_n$ *are pairwise uncorrelated, i.e.* $cov(X_i, X_j) = 0$ *for* $i, j = 1, 2, \ldots, n$ and $i \neq j$, *then*

$$var(a_1X_1 + a_2X_2 + \cdots + a_nX_n)$$
$$= a_1^2\, var(X_1) + a_2^2\, var(X_2) + \cdots + a_n^2\, var(X_n). \quad (8.4.9)$$

Remark. If $X_1, X_2, \ldots, X_n$ are mutually independent, then they are pairwise independent and so they are pairwise uncorrelated and consequently (8.4.9) holds if $X_1, X_2, \ldots, X_n$ are mutually independent.

8.5. Joint Characteristic Function.

The *joint characteristic function* of the joint distribution of the random variables X and Y is a complex valued function $\phi_{X,Y}(t, u)$ of the two real variables t, u, defined by

$$\phi_{X,Y}(t, u) = E\{e^{i(tX+uY)}\}, \quad (8.5.1)$$

where $i = \sqrt{-1}$.

It can be shown that $E[e^{i(tX+uY)}]$ exists for any distribution (discrete or continuous) for all real values of t, u. Let $\phi_X(t), \phi_Y(u)$ be respectively the characteristic functions of X and Y. Then we see that $\phi_{X,Y}(0, u) = \phi_Y(u)$ and $\phi_{X,Y}(t, 0) = \phi_X(t)$. We have proved in Theorem 8.1.1(d) that

$$E\{g_1(X)\, g_2(Y)\} = E\{g_1(X)\}\, E\{g_2(Y)\},$$

if X, Y are independent.

Now let $U = g_1(X) + ih_1(X)$

$V = g_2(Y) + ih_2(Y)$,

where $i = \sqrt{-1}$. Now we shall prove that

$$E(UV) = E(U)\,E(V), \tag{8.5.2}$$

if X, Y are independent. We see that

$$E(UV) = E[\{g_1(X) + ih_1(X)\}\{g_2(Y) + ih_2(Y)\}]$$
$$= E\{g_1(X)\,g_2(Y) - h_1(X)h_2(Y)\} + iE\{g_1(X)h_2(Y) + h_1(X)g_2(Y)\}$$
$$= [E\{g_1(X)\}\,E\{g_2(Y)\} - E\{h_1(X)\}E\{h_2(Y)\}]$$
$$+ i[E\{g_1(X)\}\,E\{h_2(Y)\} + E\{h_1(X)\}\,E\{g_2(Y)\}],$$

since X, Y are independent

$$= [E\{g_1(X)\} + iE\{h_1(X)\}][E\{g_2(Y)\} + iE\{h_2(Y)\}].$$
$$= E\{g_1(X) + ih_1(X)\}\,E\{g_2(Y) + ih_2(Y)\}$$
$$= E(U)\,E(V).$$

Hence (8.5.2) is proved.

THEOREM 8.5.1. *If X, Y are independent random variables, then*

$$\phi_{X,Y}(t, u) = \phi_X(t)\,\phi_Y(u), \tag{8.5.3}$$

with usual notations.

Proof: We have

$$\phi_{X,Y}(t, u) = E\left\{e^{i(tX+uY)}\right\} = E\left(e^{itX}\cdot e^{iuY}\right)$$
$$= E(e^{itX})\,E\,e^{iuY}) \qquad \text{by (8.5.2)}$$
$$= \phi_X(t)\,\phi_Y(u).$$

Hence the theorem.

Joint characteristic function of n random variables $X_1, X_2, \ldots, X_n$.

The joint characteristic function $\phi_{X_1, X_2, \ldots, X_n}(t_1, t_2, \ldots, t_n)$ of the joint distribution of $X_1, X_2, \ldots, X_n$, is a complex valued function of n real variables $t_1, t_2, \ldots, t_n$, defined by

$$\phi_{X_1, X_2, \ldots, X_n}(t_1, t_2, \ldots, t_n) = E\left\{e^{i(t_1X_1 + t_2X_2 + \cdots + t_nX_n)}\right\}, \tag{8.5.4}$$

where $i = \sqrt{-1}$.

Generalising the theorem 8.5.1 for n random variables we get the following theorem :

THEOREM 8.5.2. *If $X_1, X_2, \ldots, X_n$ are mutually independent random variables with characteristic functions $\phi_{X_1}(t_1), \phi_{X_2}(t_2), \ldots, \phi_{X_n}(t_n)$ respectively, then*

$$\phi_{X_1, X_2, \ldots, X_n}(t_1, t_2, \ldots, t_n) = \phi_{X_1}(t_1)\,\phi_{X_2}(t_2)\ldots\ldots\phi_{X_n}(t_n).$$

Proof: Since $X_1, X_2, \ldots, X_n$ are mutually independent, it can be shown easily that

$$E(U_1 U_2 \ldots\ldots U_n) = E(U_1)\,E(U_2)\ldots\ldots E(U_n), \tag{8.5.5}$$

where $U_r = g_r(X_r) + ih_r(X_r)$, for $r = 1, 2, \ldots, n$; $i = \sqrt{-1}$.

Then $\phi_{X_1 X_2, \ldots, X_n}(t_1, t_2, \ldots, t_n)$

$$= E(e^{it_1X_1} \cdot e^{it_2X_2} \ldots\ldots e^{it_nX_n})$$
$$= E(e^{it_1X_1})\,E(e^{it_2X_2}) \ldots\ldots E(e^{it_nX_n}) \qquad \text{by (8.5.5)}$$
$$= \phi_{X_1}(t_1)\,\phi_{X_2}(t_2)\ldots\ldots\phi_{X_n}(t_n)$$

Hence the theorem is proved.

The converse of Theorem 8.5.2 is also true. Since the proof is beyond the scope of this treatise, we just state the converse theorem below :

THEOREM 8.5.3. *If the joint characteristic function of n random variables $X_1\, X_2, \ldots, X_n$ be given by*

$\phi_{X_1, X_2, \ldots, X_n}(t_1, t_2, \ldots, t_n) = \phi_{X_1}(t_1)\,\phi_{X_2}(t_2) \ldots \phi_{X_n}(t_n)$, where $\phi_{X_1}(t_1), \phi_{X_2}(t_2), \ldots, \phi_{X_n}(t_n)$ are respectively the characteristic functions of $X_1, X_2, \ldots, X_n$, then the random variables $X_1, X_2 \ldots, X_n$ are independent.

Reproductive properties of various distributions :

A. **Binomial Distribution.**

THEOREM 8.5.4. *If $X_1, X_2, \ldots, X_r$ be mutually independent binomial variates with parameters $(n_1, p), (n_2, p) \cdots\cdots (n_r, p)$ respectively, then $X_1 + X_2 + \cdots\cdots + X_r$ is a binomial $(n_1 + n_2 + \cdots\cdots + n_r, p)$ variate.*

Proof: Let $S_r = X_1 + X_2 + \cdots\cdots + X_r$. Here the characteristic function X_k is given by

$$\phi_{X_k}(t) = E(e^{itX_k}), \ k = 1, 2, \ldots r$$
$$= (pe^{it} + q)^{n_k}.$$

$q = 1 - p$ and $i = \sqrt{-1}$.

Then the characteristic function $\phi_{S_r}(t)$ of S_r is given by

$$\phi_{S_r}(t) = E(e^{itS_r})$$
$$= E\{e^{i(tX_1 + tX_2 + \cdots\cdots + tX_r)}\}$$
$$= E(e^{itX_1} \cdot e^{itX_2} \cdots\cdots e^{itX_r}).$$

Now $X_1, X_2, \ldots, X_r$ are mutually independent random variables. Then by (8.5.5) we get

$$E(e^{itX_1} \cdot e^{itX_2} \cdot \cdots\cdots e^{itX_r})$$
$$= E(e^{itX_1}) E(e^{itX_2}) \cdots\cdots E(e^{itX_r})$$
$$= (pe^{it} + q)^{n_1} (pe^{it} + q)^{n_2} \cdots\cdots (pe^{it} + q)^{n_r}$$
$$= (pe^{it} + q)^{n_1 + n_2 + \cdots\cdots + n_r}.$$

Hence we get

$$\phi_{S_r}(t) = (pe^{it} + q)^{n_1 + n_2 + \cdots\cdots + n_r}$$

which is the characteristic function of a binomial $(n_1 + n_2 + \cdots + n_r, p)$ variate.

So, by the uniqueness theorem of characteristic function, it is proved that $S_r = X_1 + X_2 + \cdots + X_r$ is a binomial $(n_1 + n_2 + \cdots + n_r, p)$ variate.

B. Poisson Distribution.

THEOREM 8.5.5. *If $X_1, X_2, \ldots\ldots, X_r$ be mutually independent Poisson variates with parameters $\mu_1, \mu_2, \ldots\ldots, \mu_r$ respectively, then $X_1 + X_2 + \cdots\cdots + X_r$ has Poisson distribution with parameter $\mu_1 + \mu_2 + \cdots + \mu_r$.*

Proof: Here the characteristic function $\phi_{X_k}(t)$ of X_k is given by $\phi_{X_k}(t) = e^{\mu_k(e^{it} - 1)}$ for $k = 1, 2, \ldots, r$. Then the characteristic function $\phi_{S_r}(t)$ of $S_r = X_1 + X_2 + \cdots + X_r$ is given by

$$\phi_{S_r}(t) = E\{e^{it(X_1 + X_2 + \cdots + X_r)}\}$$
$$= E[e^{itX_1} e^{itX_2} \cdots\cdots e^{itX_r}]$$
$$= E(e^{itX_1}) E(e^{itX_2}) \cdots\cdots E(e^{itX_r}),$$

since $X_1, X_2, \ldots, X_r$ are mutually independent. Then we get

$$\phi_{S_r}(t) = \phi_{X_1}(t) \cdots\cdots \phi_{X_r}(t)$$
$$= e^{\mu_1(e^{it} - 1)} e^{\mu_2(e^{it} - 1)} \cdots\cdots e^{\mu_r(e^{it} - 1)}$$
$$= e^{\{\mu_1(e^{it} - 1) + \mu_2(e^{it} - 1) + \cdots\cdots + \mu_r(e^{it} - 1)\}}$$
$$= e^{(\mu_1 + \mu_2 + \cdots\cdots + \mu_r)(e^{it} - 1)}.$$

So the characteristic function of $S_r = X_1 + X_2 + \cdots\cdots + X_r$ is given by

$\phi_{S_r}(t) = e^{(\mu_1 + \mu_2 + \cdots\cdots + \mu_r)(e^{it} - 1)}$, which is the characteristic function of a Poisson $(\mu_1 + \mu_2 + \cdots\cdots + \mu_r)$ variate. Hence, by the uniqueness theorem of characteristic function, it is proved that $S_r = X_1 + X_2 + \cdots\cdots + X_r$ has Poisson distribution with parameter $\mu_1 + \mu_2 + \cdots\cdots + \mu_r$.

C Normal Distribution.

THEOREM 8.5.6. *If $X_1, X_2, \ldots, X_n$ be mutually independent normal variates with parameters $(m_1, \sigma_1), (m_2, \sigma_2), \ldots, (m_n, \sigma_n)$ respectively, then $X_1 + X_2 + \cdots\cdots + X_n$ has normal distribution with parameters $(m_1 + m_2 + \cdots\cdots + m_n, \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots\cdots + \sigma_n^2})$.*

Proof: Here the characteristic function $\phi_{X_k}(t)$ of X_k is given by $\phi_{X_k}(t) = e^{im_kt - \frac{1}{2}\sigma_k^2t^2}$ for $k = 1, 2, \ldots, n$. Then the characteristic function of $S_n = X_1 + X_2 + \cdots\cdots + X_n$ is given by

$$\phi_{S_n}(t) = E\{e^{it(X_1 + X_2 + \cdots\cdots + X_n)}\}$$

So $\phi_{S_n}(t) = E\{e^{itX_1} \cdot e^{itX_2} \cdots\cdots e^{itX_n}\}$

$$= E(e^{itX_1}) \cdot E(e^{itX_2}) \cdots\cdots E(e^{itX_n}),$$

since $X_1, X_2, \ldots, X_n$ are mutually independent. Then we get

$$\phi_{S_n}(t) = \phi_{X_1}(t)\, \phi_{X_2}(t) \cdots\cdots \phi_{X_n}(t)$$

$$= e^{im_1t - \frac{1}{2}\sigma_1^2t^2} \cdot e^{im_2t - \frac{1}{2}\sigma_2^2t^2} \cdots\cdots e^{im_nt - \frac{1}{2}\sigma_n^2t^2}$$

$$= e^{i(m_1 + m_2 + \cdots\cdots + m_n)t - \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \cdots\cdots + \sigma_n^2)t^2}.$$

So the characteristic function of $S_n = X_1 + X_2 + \cdots\cdots + X_n$ is given by

$$\phi_{S_n}(t) = e^{i(m_1 + m_2 + \cdots\cdots + m_n)t - \frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \cdots\cdots + \sigma_n^2)t^2},$$

which is the characteristic function of a normal

$(m_1 + m_2 + \cdots\cdots + m_n,\ \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots\cdots + \sigma_n^2})$ variate. So, by the Uniqueness theorem of characteristic function, it is proved that $S_n = X_1 + X_2 + \cdots\cdots + X_n$ has normal distribution with parameters $(m_1 + m_2 + \cdots + m_n,\ \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots\cdots + \sigma_n^2})$.

We now prove the following general theorem, from which we can obtain the reproductive property of normal distribution as a particular case.

THEOREM 8.5.7. *If $X_1, X_2, \ldots, X_n$ are mutually independent normal variates with parameters $(m_1, \sigma_1), (m_2, \sigma_2), \ldots, (m_n, \sigma_n)$ respectively and if $a_1, a_2, \ldots, a_n$ be any real constants, then $a_1X_1 + a_2X_2 + \cdots + a_nX_n$ has normal distribution with parameters.*

$$(a_1m_1 + a_2m_2 + \cdots + a_nm_n,\ \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2}),$$

Proof: Here the characteristic function $\phi_{X_k}(t)$ of X_k is given by $\phi_{X_k}(t) = e^{im_kt - \frac{1}{2}\sigma_k^2t^2}$ for $k = 1, 2, \ldots, n$.

Then the characteristic function of $S_n = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ is given by

$$\phi_{S_n}(t) = E\{e^{it(a_1X_1 + a_2X_2 + \cdots + a_nX_n)}\}$$

$$= E(e^{ita_1X_1} \cdot e^{ita_2X_2} \cdots e^{ita_nX_n})$$

$$= E(e^{ita_1X_1})E(e^{ita_2X_2}) \cdots E(e^{ita_nX_n}),$$

since $X_1, X_2, \ldots, X_n$ are mutually independent.

Then we get $\phi_{S_n}(t) = \phi_{X_1}(a_1t)\, \phi_{X_2}(a_2t) \cdots \phi_{X_n}(a_nt)$.

Now, $\phi_{X_k}(a_kt) = e^{ia_km_kt - \frac{1}{2}a_k^2\sigma_k^2t^2}$ for $k = 1, 2, \ldots, n$.

$$\therefore\ \phi_{S_n}(t) = e^{ia_1m_1t - \frac{1}{2}a_1^2\sigma_1^2t^2} \cdot e^{ia_2m_2t - \frac{1}{2}a_2^2\sigma_2^2t^2} \cdots e^{ia_nm_nt - \frac{1}{2}a_n^2\sigma_n^2t^2}$$

$$= e^{i(a_1m_1 + a_2m_2 + \cdots + a_nm_n)t - \frac{1}{2}(a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2)t^2}$$

So the charcteristic function of $S_n = a_1X_1 + a_2X_2 + \cdots\cdots + a_nX_n$ is given by

$$\phi_{S_n}(t) = e^{i(a_1m_1 + a_2m_2 + \cdots + a_nm_n)t - \frac{1}{2}(a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2)t^2}$$

which is the characteristic function of a normal $(a_1m_1 + a_2m_2 + \cdots + a_nm_n,\ \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2})$ variate. Hence, by the Uniqueness theorem of characteristic function, it is proved that $S_n = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ has normal distribution with parameters $(a_1m_1 + a_2m_2 + \cdots + a_nm_n,\ \sqrt{a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2})$.

D. Gamma Distribution.

THEOREM 8.5.8. *If $X_1, X_2, \ldots, X_n$ are mutually independent gamma variates with parameters $l_1, l_2, \ldots, l_n$ respectively, then $X_1 + X_2 + \cdots + X_n$ has gamma distribution with parameter $l_1 + l_2 + \cdots + l_n$.*

Proof: Here the characteristic function of X_k is given by

$$\phi_{X_k}(t) = (1 - it)^{-l_k}, \text{ for } k = 1, 2, \ldots, n.$$

Then the characteristic function $\phi_{S_n}(t)$ of $S_n = X_1 + X_2 + \cdots + X_n$ is given by

$$\phi_{S_n}(t) = E\{e^{it(X_1 + X_2 + \cdots + X_n)}\}$$

$$= E(e^{itX_1} \cdot e^{itX_2} \cdots\cdots e^{itX_n})$$

$$= E(e^{itX_1})E(e^{itX_2}) \cdots\cdots E(e^{itX_n}),$$

since $X_1, X_2, \ldots, X_n$ are mutually independent. Then we get

$$\phi_{S_n}(t) = \phi_{X_1}(t)\, \phi_{X_2}(t) \cdots\cdots \phi_{X_n}(t)$$

$$= (1 - it)^{-l_1}(1 - it)^{-l_2} \cdots\cdots (1 - it)^{-l_n}$$

$$= (1 - it)^{-(l_1 + l_2 + \cdots + l_n)}.$$

So the characteristic function of $S_n = X_1 + X_2 + \cdots + X_n$ is given by

$$\phi_{S_n}(t) = (1 - it)^{-(l_1 + l_2 + \cdots + l_n)}$$

which is the characteristic function of a gamma $(l_1 + l_2 + \cdots + l_n)$ variate. Hence by the Uniqueness theorem of characteristic function, $S_n = X_1 + X_2 + \cdots + X_n$ has gamma distribution with parameter $l_1 + l_2 + \cdots + l_n$.

8.6. Conditional Expectations, Regression curves.

We consider the joint distribution of the random variables X and Y. Let $g : R \times R \to R$ be a continuous function. The conditional expectation of $g(X, Y)$, given $X = x$, denoted as $E\{g(X, Y) \mid X = x\}$, is defined in the following cases :

Case I. Joint distribution of X, Y is discrete.

Let $f_{ij} = P(X = x_i, Y = y_j)$, where (x_i, y_j) is a point of the spectrum of (X, Y). If $x = x_i$, then

$E\{g(X, Y) \mid X = x\} = E\{g(X, Y) \mid X = x_i\}$

is defined by $E\{g(X, Y) \mid X = x_i\} = \sum_j g(x_i, y_j) \dfrac{f_{ij}}{f_{i\cdot}}$, (8.6.1)

where $f_{i\cdot} = \sum_j f_{ij}$, provided the series on the right hand side of (8.6.1) is absolutely convergent. If x is not equal to any x_i (x_i is a point of the spectrum of X) then, $E\{g(X, Y) \mid X = x\}$ cannot be defined.

Case II. Joint distribution of (X, Y) is continuous.

Let $f(x, y)$ be the joint density function of X and Y.

Here, for a fixed x, $E\{g(X, Y) \mid X = x\}$ is defined by

$$E\{g(X, Y) \mid X = x\} = \int_{-\infty}^{\infty} g(x, y) f_{Y \mid X}(y \mid x)\, dy. \tag{8.6.2}$$

where $f_{Y \mid X}(y \mid x) = \dfrac{f(x, y)}{f_X(x)}$ and $f_X(x) = \int_{-\infty}^{\infty} f(x, t)\, dt$,

provided the integral in the right hand side of (8.6.2) is absolutely convergent.

The expressions for $E\{g(X, Y) \mid X = x\}$ given in (8.6.1) and (8.6.2) indicate that the '*conditional expectation* of $g(X, Y)$, given $X = x$' is nothing but the expectation of the random variable $g(x, Y) = \varphi(Y)$ (say) with respect to the conditional distribution of Y on the hypothesis $X = x$.

We can similarly define $E\{g(X, Y) \mid Y = y\}$ in the above two cases. Thus we have

$$E\{g(X, Y) \mid Y = y_j\} = \sum_i g(x_i, y_j) \frac{f_{ij}}{f_{\cdot j}}, \tag{8.6.3}$$

if (X, Y) is discrete, where $f_{\cdot j} = \sum_i f_{ij}$ and for a fixed y_j.

$$E\{g(X, Y) \mid Y = y\} = \int_{-\infty}^{\infty} g(x, y) \frac{f(x, y)}{f_Y(y)}\, dx, \tag{8.6.4}$$

if (X, Y) is continuous, where $f_Y(y) = \int_{-\infty}^{\infty} f(t, y)\, dt$, provided the series in (8.6.3) and the integral in (8.6.4) are absolutely convergent. As usual we say that $E\{g(X, Y) \mid X = x\}$ and $E\{g(X, Y) \mid Y = y\}$ exist if the corresponding series or integral defining them are absolutely convergent.

Conditional Moments, Conditional Means and Conditional Variances.

Let r be a given positive integer. If $E(Y^r \mid X = x)$ exists, then $E(Y^r \mid X = x)$ is called the *r-th order conditional moment of* Y about the origin on the hypothesis $X = x$. Likewise $E(X^r \mid Y = y)$, (if it exists), is called the *r-th order conditional moment of* X about the origin on the hypothesis $Y = y$.

Let us consider the particular case $r = 1$ in the following cases :

Case I. Joint distribution of X, Y is discrete.

The conditional expectation $E(Y \mid X = x_i)$, if it exists, is called the conditional mean of Y on the hypothesis $X = x_i$ and it is denoted by $m_{Y \mid i}$; so $m_{Y \mid i}$ is given by

$$m_{Y \mid i} = \frac{\sum_j y_j f_{ij}}{f_{i\cdot}}. \tag{8.6.5}$$

where (x_i, y_j), f_{ij}, $f_{i\cdot}$, have the usual meanings. Similarly, the conditional mean of X on the hypothesis $Y = y_j$, denoted by $m_{X \mid j}$ is given by (when it exists)

$$m_{X \mid j} = \frac{\sum_i x_i f_{ij}}{f_{\cdot j}}. \tag{8.6.6}$$

Case II. Distribution of (X, Y) is continuous.

The conditional expectation $E(Y \mid X = x)$, if it exists, is called the conditional mean of Y on the hypothesis $X = x$ and it is denoted by $m_Y(x)$ and so $m_Y(x)$ is given by

$$m_Y(x) = \frac{\int_{-\infty}^{\infty} y f(x, y)\, dy}{f_X(x)} \tag{8.6.7}$$

where $f(x, y)$, $f_X(x)$ have the usual meanings.

Similarly the conditional mean of X on the hypothesis $Y=y$, denoted by $m_X'(y)$, is given by

$$m_X(y)=\frac{\int_{-\infty}^{\infty} x f(x, y)\,dx}{f_Y'(y)} \tag{8.6.8}$$

The conditional expectation $E\{(Y-m_{Y|i})^2 \mid X=x_i\}$, if it exists, when (X, Y) is discrete, is called the *conditional variance* of Y on the hypothesis $X=x_i$ and it is denoted by $\sigma^2_{Y|i}$ where the non-negative number $\sigma_{Y|i}$ is called the *conditional standard deviation* of Y on the hypothesis $X=x_i$. Similarly $\sigma_{X|j}$ can be defined when (X, Y) is discrete. Thus when (X, Y) is discrete, $\sigma_{Y|i}$ and $\sigma_{X|j}$ (if they exist) are given by

$$\sigma^2_{Y|i}=\sum_j (y_j-m_{Y|i})^2 \frac{f_{ij}}{f_{i\cdot}} \tag{8.6.9}$$

$$\sigma^2_{X|j}=\sum_i (x_i-m_{X|j})^2 \frac{f_{ij}}{f_{\cdot j}} \tag{8.6.10}$$

The conditional expectation $E[\{Y-m_Y(x)\}^2 \mid X=x]$, if it exists, when (X, Y) is continuous, is called the *conditional variance of* Y on the hypothesis $X=x$ and it is denoted by $\sigma_Y^2(x)$. Similarly we can define $\sigma_X^2(y)$ when (X, Y) is continuous.

Thus when (X, Y) is continuous $\sigma^2_Y(x)$ and $\sigma_X^2(y)$ (if they exist are given by

$$\sigma^2_Y(x)=\int_{-\infty}^{\infty} \{y-m_Y(x)\}^2 \frac{f(x, y)}{f_X(x)}\,dy \tag{8.6.11}$$

$$\sigma^2_X(y)=\int_{-\infty}^{\infty} \{x-m_X(y)\}^2 \frac{f(x, y)}{f_Y(y)}\,dx. \tag{8.6.12}$$

Significance of Conditional Means.

The expressions for the conditional means given by (8.6.5), (8.6.6) in the discrete case indicate that the conditional means $m_{Y|i}$, $m_{X|j}$ give the positions of the centre of mass of the bivariate probability mass particles situated respectively on the lines $x=x_i$ and $y=y_j$.

Also the expressions of the conditional means given by (8.6.7) and (8.6.8) in the continuous case indicate that the conditional means $m_Y'(x)$, $m_X'(y)$ give the positions of the centre of mass of the bivariate probability mass respectively on the infinitesimal strip (parallel to y-axis) between x, $x+dx$ and on the infinitesimal strip (parallel to x-axis) between y, $y+dy$.

Meaning of E{g(Y)|X} and E{g(X)|Y}.

The value of conditional expectation $E\{g(Y) \mid X=x\}$ when it exists, varies as x varies and then $E\{g(Y) \mid X=x\}$ defines a real valued function of a real variable x and let us denote this function by $h(x)$. Similarly we can define the function $l(y)$ by

$$l(y)=E\{g(X) \mid Y=y\}$$

provided the conditional expectation exists. The random variables corresponding to the real valued functions $h(x), l(y)$ are respectively $h(X), l(Y)$. Then the random variables $h(X)$, $l(Y)$ are respectively denoted as $E\{g(Y) \mid X\}$ and $E\{g(X) \mid Y\}$. Now if $E\{h(X)\}$ and $E\{l(Y)\}$ exist, then

$$E[E\{g(Y) \mid X\}]=E\{h(X)\}=\int_{-\infty}^{\infty} h(x)\, f_X(x)\,dx$$
$$=\int_{-\infty}^{\infty} E\{g(Y) \mid X=x\}\, f_X(x)\,dx \tag{8.6.13}$$

and

$$E[E\{g(X) \mid Y\}]=E\{l(Y)\}=\int_{-\infty}^{\infty} l(y)\, f_Y(y)\,dy$$
$$=\int_{-\infty}^{\infty} E\{g(X) \mid Y=y\}\, f_Y(y)\,dy \tag{8.6.14}$$

when (X, Y) is continuous. Similarly we can obtain expressions for $E[E\{g(Y) \mid X\}]$ and $E[E\{g(X) \mid Y\}]$ when (X, Y) is discrete.

THEOREM 8.6.1. *If (X, Y) be a two-dimensional random variable and $g_1 : R \to R$ and $g_2 : R \to R$ be two continuous functions, then*

(a) $E[\{g_1(Y)+g_2(Y)\} \mid X=x]$
$$=E\{g_1(Y) \mid X=x\}+E\{g_2(Y) \mid X=x\}, \tag{8.6.15}$$

(b) $$E\{g_1(Y)\, g_2(X) \mid X=x\}=g_2(x)\, E\{g_1(Y) \mid X=x\} \tag{8.6.16}$$

provided the conditional expectations in the right hand sides of (8.6.15) *and* (8.6.16) *exist.*

Proof: (a) Here $E\{g_1(Y) \mid X=x\}$ and $E\{g_2(Y) \mid X=x\}$ exist. We first consider the case when (X, Y) is continuous.

Then $\int_{-\infty}^{\infty} g_1(y) \frac{f(x, y)}{f_X(x)}\,dy$, $\int_{-\infty}^{\infty} g_2(y) \frac{f(x, y)}{f_X(x)}\,dy$ are absolutely convergent, where $f(x, y)$, $f_X(x)$ have the usual meanings.

Hence, $\int_{-\infty}^{\infty}\{g_1(y)+g_2(y)\}\dfrac{f(x,y)}{f_X(x)}\,dy$ is absolutely convergent and

$$\int_{-\infty}^{\infty}\{g_1(y)+g_2(y)\}\frac{f(x,y)}{f_X(x)}\,dy$$

$$=\int_{-\infty}^{\infty}g_1(y)\frac{f(x,y)}{f_X(x)}\,dy+\int_{-\infty}^{\infty}g_2(y)\frac{f(x,y)}{f_X(x)}\,dy.$$

But, $\int_{-\infty}^{\infty}g_1(y)\dfrac{f(x,y)}{f_X(x)}\,dy=E\{g_1(Y)\,|\,X=x\}$ and

$$\int_{-\infty}^{\infty}g_2(y)\frac{f(x,y)}{f_X(x)}\,dy=E\{g_2(Y)\,|\,X=x\}.$$

So $E[\{g_1(Y)+g_2(Y)\}\,|\,X=x]$ exists and

$$E[\{g_1(Y)+g_2(Y)\}\,|\,X=x]=E\{g_1(Y)\,|\,X=x\}+E\{g_2(Y)\,|\,X=x\}.$$

So *(a)* is proved when (X,Y) is continuous. Similarly we can prove *(a)* when (X,Y) is discrete and where $x=x_i$, x_i being a point of the spectrum of X.

(b) Here $E\{g_1(Y)\,|\,X=x\}$ exists. Then in the continuous case $\int_{-\infty}^{\infty}g_1(y)\dfrac{f(x,y)}{f_X(x)}\,dy$ is absolutely convergent and hence

$$\int_{-\infty}^{\infty}g_2(x)\,g_1(y)\,\frac{f(x,y)}{f_X(x)}\,dy$$

is absolutely convergent and consequently

$$\int_{-\infty}^{\infty}g_2(x)\,g_1(y)\,\frac{f(x,y)}{f_X(x)}\,dy=E\{g_1(Y)\,g_2(X)\,|\,X=x\}.$$

So we get

$$E\{g_1(Y)\,g_2(X)\,|\,X=x\}=g_2(x)\int_{-\infty}^{\infty}g_1(y)\,\frac{f(x,y)}{f_X(x)}\,dy$$

$$=g_2(x)\,E\{g_1(Y)\,|\,X=x\}.$$

Hence *(b)* is proved when (X,Y) is continuous. Similarly we can prove *(b)* when (X,Y) is discrete and where $x=x_i$, x_i being a point of the spectrum of X.

THEOREM 8.6.2. *If $g:R\to R$ be a continuous function and if $E\{g(Y)\}$ exists, then*

$$E[E\{g(Y)\,|\,X\}]=E\{g(Y)\}. \qquad (8.6.17)$$

Proof: Case I. Let the distribution of (X,Y) be continuous.

We have $E\{g(Y)\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(y)\,f(x,y)\,dx\,dy,$

where $f(x,y)$ is the probability density function of (X,Y)

$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}g(y)\,f(x,y)\,dy\right\}dx$$

$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}g(y)\,\frac{f(x,y)}{f_X(x)}\,dy\right\}f_X(x)\,dx,$$

$f_X(x)$ being the probability density function of X

$$=\int_{-\infty}^{\infty}E\{g(Y)\,|\,X=x\}\,f_X(x)\,dx$$

$$=E[E\{g(Y)\,|\,X\}].$$

Hence, the theorem is proved when (X,Y) is continuous.

Case II. Let the distribution of (X,Y) be discrete.

Here $E\{g(Y)\}=\sum_i\sum_j g(y_j)\,f_{ij},$

where $P(X=x_i,\,Y=y_j)=f_{ij}$, (x_i,y_j) being a spectrum point of (X,Y).

$$\therefore\quad E\{g(Y)\}=\sum_i\left\{\sum_j g(y_j)\,\frac{f_{ij}}{f_{i\cdot}}\right\}f_{i\cdot},\ \text{where } f_{i\cdot}=\sum_j f_{ij}$$

$$=\sum_i h(x_i)\,f_{i\cdot},\ \text{where } h(x_i)=\sum_j g(y_j)\,\frac{f_{ij}}{f_{i\cdot}}=E\{g(Y)\,|\,X=x_i\}$$

and the random variable $h(X)=E\{g(Y)\,|\,X\}$

$$=E\{h(X)\}=E[E\{g(Y)\,|\,X\}].$$

Hence $E[E\{g(Y)\,|\,X\}]=E\{g(Y)\}$, when (X,Y) is discrete.

Cor. $E\{E(Y\,|\,X)\}=E(Y). \qquad (8.6.18)$

Taking $g(Y)=Y$ in (8.6.17), we get $E\{E(Y\,|\,X)\}=E(Y)$.

THEOREM 8.6.3. *If (X,Y) be continuous and for a given x, the conditional variance $\sigma_Y^2(x)$ exists, then*

$$\sigma_Y^2(x)=E(Y^2\,|\,X=x)-\{m_Y(x)\}^2. \qquad (8.6.19)$$

Proof: Here $\sigma_Y^2(x)=E[\{Y-m_Y(x)\}^2\,|\,X=x]$

$$=E[\{Y^2-2m_Y(x)\,Y+m_Y^2(x)\}\,|\,X=x]$$

$$=E(Y^2\,|\,X=x)+E[\{m_Y^2(x)-2m_Y(x)Y\}\,|\,X=x]$$

by *(a)* of Theorem 8.6.1.

Again we have

$E[\{m_Y^2(x) - 2m_Y(x)Y\} \mid X = x]$

$= E\{m_Y^2(x) \mid X = x\} + E\{-2m_Y(x)Y \mid X = x\}$, by (a) of Theorem 8.6.1

$$= \int_{-\infty}^{\infty} m_Y^2(x)\frac{f(x, y)}{f_X(x)}\,dy + \int_{-\infty}^{\infty} (-2)\,m_Y(x)\,y\,\frac{f(x, y)}{f_X(x)}\,dy$$

$= m_Y^2(x) - 2m_Y(x)\,E(Y \mid X = x)$

$= m_Y^2(x) - 2m_Y(x)\,m_Y(x)$

$= -m_Y^2(x)$.

Hence, we get

$$\sigma_Y^2(x) = E(Y^2 \mid X = x) - \{m_Y(x)\}^2.$$

Note. It can be proved similarly that

$$(\sigma_{Y \mid i})^2 = E(Y^2 \mid X = x_i) - (m_{Y \mid i})^2, \tag{8.6.20}$$

when (X, Y) is discrete.

8·7. Regression Curves.

Let the joint distribution of X and Y be continuous. If $m_Y(x)$ exists, where $x \in R$, then $m_Y(x)$ gives a real valued function of a real variable x, defined in some domain, say D, where $D \subseteq R$ and this function is called the regression function of Y on X. In this case the equation $y = m_Y(x)$ defines a curve which is called the regression curve of Y on X.

Similarly if $m_X(y)$ exists $(y \in R)$, then the function defined by $m_X(y)$ is called the regression function of X on Y and the curve $x = m_X(y)$ is called the regression curve of X on Y

The curves $y = m_Y(x)$ and $x = m_X(y)$ are also called '*regression curve for the mean of Y*' and '*regression curve for the mean of X*' respectively.

The following two theorems give geometrical interpretations of the regression curves for a bivariate continuous distribution.

THEOREM 8.7.1. *For a continuous two-dimensional variate (X, Y) and for any continuous function $g(x)$, $E[\{Y - g(X)\}^2]$ is minimum when $g(x) = m_Y(x)$.*

Proof: Let $f(x, y)$ be the joint probability density function of of X and Y. Then $E[\{Y - g(X)\}^2]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{y - g(x)\}^2 f(x, y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \{y - g(x)\}^2 f(x, y)\,dy\right]dx$$

$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} \{y - g(x)\}^2 f_Y(y \mid x) f_X(x)\,dy\right]dx$$

$\left[\because \dfrac{f(x, y)}{f_X(x)} = f_Y(y \mid x)\right.$, which is the conditional density function of Y on the hypothesis $X = x$]

$$= \int_{-\infty}^{\infty} f_X(x)\left[\int_{-\infty}^{\infty} \{y - g(x)\}^2 f_Y(y \mid x)\,dy\right]dx$$

$$= \int_{-\infty}^{\infty} f_X(x) E[\{Y - g(X)\}^2 \mid X = x]\,dx.$$

Now $E[\{Y - g(X)\}^2 \mid X = x]$

$= E[\{Y - m_Y(X) + m_Y(X) - g(X)\}^2 \mid X = x]$

$= E[\{Y - m_Y(X)\}^2 \mid X = x] + \{m_Y(x) - g(x)\}^2$
$\qquad + 2\{m_Y(x) - g(x)\}E[\{Y - m_Y(X)\} \mid X = x]$.

Again $E[\{Y - m_Y(X)\} \mid X = x]$

$$= \int_{-\infty}^{\infty} \{y - m_Y(x)\}\frac{f(x, y)}{f_X(x)}\,dy$$

$$= \int_{-\infty}^{\infty} y\,f_Y(y \mid x)\,dy - m_Y(x)\int_{-\infty}^{\infty} f_Y(y \mid x)\,dy$$

$= m_Y(x) - m_Y(x) = 0.$

So $E[\{Y - g(X)\}^2]$

$$= \int_{-\infty}^{\infty} f_X(x)\,[E\{Y - m_Y(x)\}^2 \mid X = x]\,dx + \int_{-\infty}^{\infty} \{m_Y(x) - g(x)\}^2 f_X(x)\,dx. \tag{8.7.1}$$

Now $\{m_Y(x) - g(x)\}^2 f_X(x) \geqslant 0$ for all $x \in (-\infty, \infty)$.

Also $\{m_Y(x) - g(x)\}^2 f_X(x)$ is continuous at any point x where $f_X(x)$ is continuous. So

$$\int_{-\infty}^{\infty} \{m_Y(x) - g(x)\}^2 f_X(x)\,dx \geqslant 0 \text{ and}$$

$$\int_{-\infty}^{\infty} \{m_Y(x) - g(x)\}^2 f_X(x)\,dx = 0$$ if we choose $g(x) = m_Y(x)$, where $f_X(x) > 0$.

So from (8.7.1) we find that

$$E\{Y-g(X)\}^2] > \int_{-\infty}^{\infty} f_X'(x)\, E\{Y-m_Y(X)\}^2 \mid X=x]\, dx \quad (8.7.2)$$

and the equality sign in (8.7.2) occurs if we choose $g(x)=m_Y(x)$.

Hence, it is proved that $E[\{Y-g(X)\}^2]$ is minimum if $g(X)=m_Y(X)$, *i.e.*, if $g(x)=m_Y(x)$

THEOREM 8.7.2. *For a continuous two-dimensional variate (X, Y) and for any continuous function $h(y)$, $E[\{X-h(Y)\}^2]$ is minimum when $h(y)=m_X(y)$.*

Proof: Similar to the proof of Theorem 8.7.1 and left as an exercise.

Geometrical Interpretation of Regression Curves:

Let $S_1=E[\{Y-g(X)\}^2]$ and $S_2=E[\{X-h(Y)\}^2]$ where $g(x)$, $h(y)$ are continuous functions. Now S_1 represents the mean value of the square of $Y-g(X)$ which is the random variable corresponding to the deviation of the random point (X, Y) from the curve $y=g(x)$ measured

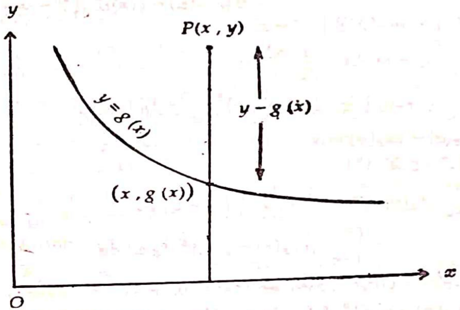


Fig. 8.7.1

parallel to the y-axis and similarly S_2 represents the mean value of the square of the random variable corresponding to the deviation of the random point (X, Y) from the curve $x=h(y)$ measured parallel to the x-axis. The aforesaid deviations are illustrated in figures 8.7.1 and 8.7.2.

Then the theorem 8.7.1 reveals that among all continuous curves $y=g(x)$, S_1 is minimum for the regression curve $y=m_Y(x)$ and the

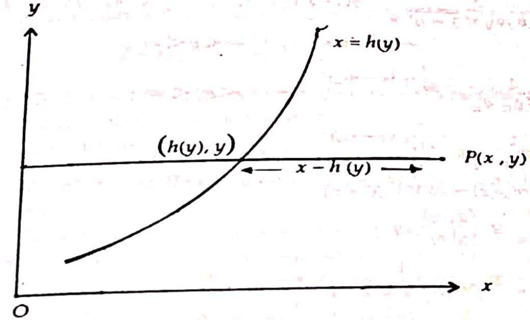


Fig. 8.7.2

theorem 8.7.2 reveals that among all continuous curves $x=h(y)$, S_2 is minimum for the regression curve $x=m_X(y)$.

THEOREM 8.7.3. *If X, Y are independent continuous variates, then the regression curves are straight lines parallel to the axes of co-ordinates.*

Proof: Here $m_Y(x)=\int_{-\infty}^{\infty} y\frac{f(x,y)}{f_X'(x)}dy$

$$=\int_{-\infty}^{\infty} y\frac{f_X(x)\, f_Y(y)}{f_X(x)}\, dy,$$

since X, Y are independent

$$=\int_{-\infty}^{\infty} y\, f_Y(y)\, dy=E(Y)=m_y.$$

Similarly, $m_X(y)=E(X)=m_x$.

So the regression curves are $y=m_y$, $x=m_x$ which are straight lines parallel to the axes of co-ordinates.

THEOREM 8.7.4. *The regression curves of a bivariate normal distribution are linear.*

Proof: Let the joint distribution of X and Y be a bivariate normal distribution with parameters m_x, m_y, σ_x, σ_y, ρ. Then the

joint probability density function $f(x, y)$ and the marginal probability density functions $f_X(x), f_Y(y)$ are given by

$$f(x, y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right)+\left(\frac{y-m_y}{\sigma_y}\right)^2\right\}}$$

$$-\infty<x<\infty,\ -\infty<y<\infty\ ;$$

$$f_X(x)=\frac{1}{\sqrt{2\pi}\,\sigma_x}\, e^{-\frac{1}{2\sigma_x^2}(x-m_x)^2},\quad -\infty<x<\infty\ ;$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\,\sigma_y}\, e^{-\frac{1}{2\sigma_y^2}(y-m_y)^2},\quad -\infty<y<\infty.$$

Now $m_Y(x)=E(Y\mid X=x)$

$$=\int_{-\infty}^{\infty} y\,\frac{f(x, y)}{f_X(x)}\,dy$$

$$=\int_{-\infty}^{\infty} y\,\frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}}\times e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right)+\left(\frac{y-m_y}{\sigma_y}\right)^2\right\}+\frac{1}{2}\left(\frac{x-m_x}{\sigma_x}\right)^2}\,dy$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} y\times e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y-m_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-m_x}{\sigma_x}\right)\left(\frac{y-m_y}{\sigma_y}\right)+\rho^2\left(\frac{x-m_x}{\sigma_x}\right)^2\right\}}\,dy$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} y\, e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-m_y}{\sigma_y}-\rho\,\frac{x-m_x}{\sigma_x}\right)^2}\,dy$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} y\, e^{\frac{-1}{2\sigma_y^2(1-\rho^2)}\left[y-\left\{m_y+\frac{\rho\sigma_y}{\sigma_x}(x-m_x)\right\}\right]^2}\,dy$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left[\sigma_y\sqrt{1-\rho^2}\,z+\left\{m_y+\frac{\rho\sigma_y}{\sigma_x}(x-m_x)\right\}\right]e^{-\frac{z^2}{2}}\,dz,$$

where $z=\frac{1}{\sigma_y\sqrt{1-\rho^2}}\left[y-\left\{m_y+\frac{\rho\sigma_y}{\sigma_x}(x-m_x)\right\}\right]$,

$$dy=\sigma_y\sqrt{1-\rho^2}\,dz$$

$$=\frac{\sigma_y\sqrt{1-\rho^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z\,e^{-\frac{z^2}{2}}\,dz+\left\{m_y+\frac{\rho\sigma_y}{\sigma_x}(x-m_x)\right\}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz$$

$$=m_y+\frac{\rho\sigma_y}{\sigma_x}(x-m_x),$$

since $\int_{-\infty}^{\infty} z\,e^{-\frac{z^2}{2}}\,dz=0$ and $\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz=\sqrt{2\pi}$.

$\therefore$ the regression curve of Y on X is

$$y=m_y+\frac{\rho\sigma_y}{\sigma_x}(x-m_x). \tag{8.7.3}$$

Similarly, the regression curve of X on Y is

$$x=m_x+\frac{\rho\sigma_x}{\sigma_y}(y-m_y). \tag{8.7.4}$$

From (8.7.3) and (8.7.4), we see that the regression curves of a bivariate normal distribution are linear.

Hence the theorem.

8.8. Principle of Least Squares—Regression Lines and Regression Parabolas.

In Theorem 8.7.1, we have proved that for the joint distribution of X and Y (X, Y are continuous variates) and for a continuous function $y=g(x)$, $E[\{Y-g(X)\}^2]$ is minimum when $g(x)=m_Y(x)$. In this case we say that from the family of all curves of the form $y=g(x)$ [$g(x)$ is continuous], the regression curve $y=m_Y(x)$ is the best fitting curve to the joint distribution of X and Y, *i.e.*, among all possible functional relationship expressed by continuous functions, the most probable relation between X and Y is given by the continuous function $y=m_Y(x)$.

In general we may be interested in finding an approximate relationship between X and Y (joint distribution of X, Y may be discrete or continuous) with the help of a curve of the form

$$y=g(x\ ;\ c_1, c_2, \ldots, c_k) \tag{8.8.1}$$

where $g(x\,;c_1, c_2, \ldots, c_k)$ is continuous and it has a known functional form and where $c_1, c_2, \ldots, c_k$ are parameters. In that case we are to find the best fitting curve to the joint distribution of X and Y from the family of curves represented by (8.8.1) and this curve is obtained by a principle known as the *principle of least squares* which consists in finding $c_1, c_2, \ldots, c_k$ for which $E[\{Y-g(X;\ c_1, c_2, \ldots, c_k\}^2]$ is minimum.

Similarly, the equation of the best fitting curve to the joint distribution of X and Y from the family of curves represented by

$$x=h(y\ ;\ d_1, d_2, \ldots, d_s) \tag{8.8.2}$$

(h is a continuous function and has a known functional form and $d_1, d_2, \ldots, d_s$ are parameters) can be obtained by the principle of

least squares, which consists in finding $d_1, d_2, \ldots, d_s$ for which $E[\{X-h(Y; d_1, d_2, \ldots, d_s)\}^2]$ is minimum.

If the family of curves represented by (8.8.1) be straight lines $y=c_1+c_2x$, then the corresponding best fitting line is called the *least square regression line* or the regression line of Y on X. Similarly the least square regression line or the regression line of X on Y is defined as the best fitting line to the joint distribution of X and Y from the family of straight lines $x=d_1+d_2y$.

If the best fitting curve be obtained from the family of curves given by $y=c_1+c_2x+c_3x^2+\cdots+c_{k+1}x^k$, then the corresponding best fitting curve is called the kth degree *least square regression parabola* or the kth degree regression parabola of Y on X and similarly we can define the kth degree regression parabola of X on Y.

Equations of the regression lines for a bivariate distribution:

Let (X, Y) be a two-dimensional variate (discrete or continuous). To find the equation of the regression line of Y on X, we are to find the values of c_1, c_2 such that $E\{(Y-c_1-c_2X)^2\}$ is minimum. Let $E(X)$, $E(Y)$, $E(X^2)$, $E(Y^2)$, $E(XY)$ exist.

Let $S=E\{(Y-c_1-c_2X)^2\}$, which exist for all values of c_1, c_2, then S can be regarded as a function of two variables c_1, c_2, where S and its partial derivatives with respect to c_1, c_2 of any order are continuous. Now a necessary condition for S to be minimum is

$$\frac{\partial S}{\partial c_1}=0, \quad \frac{\partial S}{\partial c_2}=0.$$

Now we note that

$$S=E[\{Y-(c_1+c_2X)\}^2]=\sum_i\sum_j\{y_i-(c_1+c_2x_i)\}^2 f_{ij}, \quad \text{if } (X, Y) \text{ is discrete} \tag{8.8.3}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{y-(c_1+c_2x)\}^2 f(x, y)\, dx\, dy, \quad \text{if } (X, Y) \text{ is continuous,} \tag{8.8.4}$$

where the symbols f_{ij}, $f(x, y)$ have the usual meanings. Since the series (8.8.3) and the integral (8.8.4) are absolutely convergent, the process of partial differentiation with respect to c_1, c_2 within the summation or within the integral is valid and so in either case we get

$$\frac{\partial S}{\partial c_1}=E[-2\{Y-(c_1+c_2X)\}]$$

and $$\frac{\partial S}{\partial c_2}=E[-2\{Y-(c_1+c_2X)\}X].$$

Then the equations $\dfrac{\partial S}{\partial c_1}=0$, $\dfrac{\partial S}{\partial c_2}=0$, called *normal equations*, give

$$E(Y)-c_1-c_2E(X)=0, \tag{8.8.5}$$

$$E(XY)-c_1E(X)-c_2E(X^2)=0, \tag{8.8.6}$$

or,

$$c_1+c_2m_x=m_y, \tag{8.8.7}$$

$$c_1m_x+c_2\alpha_{20}=\alpha_{11}. \tag{8.8.8}$$

Solving (8.8.7) and (8.8.8), we get

$$c_1=m_y-\frac{\operatorname{cov}(X, Y)}{\sigma_x^2}m_x=m_y-\rho\frac{\sigma_y}{\sigma_x}m_x=c_1{*}\text{(say)}, \tag{8.8.8.a}$$

$$c_2=\frac{\operatorname{cov}(X, Y)}{\sigma_x^2}=\rho\frac{\sigma_y}{\sigma_x}=c_2{*}\text{ (say)}, \tag{8.8.8 b}$$

where $\rho=\rho(X, Y)$.

Now to prove that S is minimum for $c_1=c_1*$, $c_2=c_2*$, we shall prove that $S \geqslant S*$ for all c_1, c_2, where

$$S*=E[\{Y-(c_1*+c_2*X)\}^2].$$

We have

$$\begin{aligned} S &= E[\{Y-(c_1+c_2X)\}^2] \\ &= E\{(Y-c_1*-c_2*X+c_1*+c_2*X-c_2X-c_1)^2\} \\ &= E[\{Y-(c_1*+c_2*X)+(c_1*-c_1)+(c_2*-c_2)X\}^2] \\ &= E[\{Y-(c_1*+c_2*X)\}^2]+(c_1*-c_1)^2+(c_2*-c_2)^2\times E(X^2)+2(c_1*-c_1)E(Y-c_1*-c_2*X) \\ &\quad +2(c_2*-c_2)E\{X(Y-c_1*-c_2*X)\} \\ &\quad +2(c_1*-c_1)(c_2*-c_2)E(X). \end{aligned}$$

Now by virtue of (8.8.5) and (8.8.6) we have

$$E(Y-c_1*-c_2*X)=0, \quad E\{X(Y-c_1*-c_2*X)\}=0.$$

So we get

$$\begin{aligned} S &= S*+(c_1*-c_1)^2+(c_2*-c_2)^2\alpha_{20}+2(c_1*-c_1)(c_2*-c_2)m_x \\ &= S*+(c_1*-c_1)^2+(c_2*-c_2)^2(\sigma_x^2+m_x^2) \\ &\quad +2(c_1*-c_1)(c_2*-c_2)m_x \\ &= S*+\{(c_1*-c_1)+(c_2*-c_2)m_x\}^2+(c_2*-c_2)^2\sigma_x^2. \end{aligned}$$

Now $\{(c_1*-c_1)+(c_2*-c_2)m_x\}^2+(c_2*-c_2)^2\sigma_x^2 \geqslant 0$ for all values of c_1, c_2. Hence $S \geqslant S*$ for all c_1, c_2.

Thus it is proved that S is minimum, when $c_1=c_1*$, $c_2=c_2*$.

So the equation of the regression line of Y on X is

$$y=c_1^*+c_2^*x$$

$$\text{or,}\quad y=m_y-\rho\frac{\sigma_y}{\sigma_x}m_x+\rho\frac{\sigma_y}{\sigma_x}x$$

$$\text{or,}\quad y-m_y=\rho\frac{\sigma_y}{\sigma_x}(x-m_x). \tag{8.8.9}$$

It can be proved similarly by minimizing

$$S_1=E[\{X-(d_1+d_2Y)\}^2]$$

that the equation of the regression line of X on Y is

$$x-m_x=\rho\frac{\sigma_x}{\sigma_y}(y-m_y). \tag{8.8.10}$$

Note. 1 From the equations (8.8.9) and (8.8.10) we find that the two regression lines (if they exist) intersect at (m_x, m_y).

Note. 2. We observe that the regression lines (8.8.9) and (8.8.10) coincide if $\rho=\pm 1$ and for $\rho=0$, the equations of the regression lines become $y=m_y$, $x=m_x$ which are perpendicular.

Note. 3. We have proved before that the regression curves $y=m_Y(x)$, $x=m_X(y)$ are the best fitting curves to the joint distribution of X and Y from the family of all continuous curves of the forms $y=g(x)$, $x=h(y)$ respectively. So if for a given bivariate distribution it happens that the regression curves $y=m_Y(x)$, $x=m_X(y)$ are straight lines, then the least square regression lines must coincide with the regression curves $y=m_Y(x)$, $x=m_X(y)$. In Theorem 8.7.4 we have proved that for a bivariate normal distribution the regression curves are straight lines whose equations are given by (8.7.3) and (8.7.4). So the equations of the least square regression lines for a bivariate normal distribution with parameters $m_x, m_y, \sigma_x, \sigma_y, \rho$ will be given by (8.7.3) and (8.7.4).

A measure of goodness of fit of the regression lines to the joint distribution of X and Y.

By '*goodness of fit* of a given curve $y=g(x)$ to the joint distribution of X and Y' we want to mean the goodness of approximation when we use the approximate relation $Y=g(X)$ between X and Y. Now the expression $S^*=E[\{Y-(c_1^*+c_2^*X)\}^2]$ gives the mean value of the square of deviation of the random point (X, Y) from the regression line

$$y=c_1^*+c_2^*x,\quad \left(\text{where } c_1^*=m_y-\rho\frac{\sigma_y}{\sigma_x}m_x,\ c_2^*=\rho\frac{\sigma_y}{\sigma_x}\right)$$

measured parallel to y-axis. So the goodness of fit of $y=c_1^*+c_2^*x$ will be high if the value of S^* be low and consequently S^* will give an inverse measure of goodness of fit of the regression line $y=c_1^*+c_2^*x$ to the distribution of (X, Y).

Similarly, $S_1^*=E[\{X-(d_1^*+d_2^*Y)\}^2]$ gives an inverse measure of goodness of fit of the regression line $x=d_1^*+d_2^*y$,

$$\left(\text{where } d_1^*=m_x-\rho\frac{\sigma_x}{\sigma_y}m_y,\ d_2^*=\rho\frac{\sigma_x}{\sigma_y}\right).$$

$$\text{Now } S^*=E[\{Y-(m_y-\rho\frac{\sigma_y}{\sigma_x}m_x+\rho\frac{\sigma_y}{\sigma_x}X)\}^2]$$

$$=E[\{(Y-m_y)-\rho\frac{\sigma_y}{\sigma_x}(X-m_x)\}^2]$$

$$=E\{(Y-m_y)^2\}+\rho^2\frac{\sigma_y^2}{\sigma_x^2}E\{(X-m_x)^2\}$$

$$-2\rho\frac{\sigma_y}{\sigma_x}E\{(X-m_x)(Y-m_y)\}$$

$$=\sigma_y^2+\rho^2\frac{\sigma_y^2}{\sigma_x^2}\sigma_x^2-2\rho\frac{\sigma_y}{\sigma_x}\sigma_x\sigma_y\rho$$

$$=\sigma_y^2(1-\rho^2).$$

Thus we get $S^*=\sigma_y^2(1-\rho^2)$. ... ... (8.8.11)

Similarly, we find that

$$S_1^*=\sigma_x^2(1-\rho^2). \quad \dots \quad \dots \tag{8.8.12}$$

From (8.8.11) and (8.8.12) we observe that for given values of σ_x, σ_y the values of S^* and S_1^* become low if the value of $|\rho|$ is high and for low values of $|\rho|$ we get high values of S^* and S_1^* and consequently goodness of fit of the regression lines to the bivariate distribution of (X, Y), that is, the concentration of the bivariate probability mass near the regression lines will be high or low according as the value of $|\rho|$ is high or low. So $|\rho|$ is a direct measure of goodness of fit of the regression lines to the joint distribution of X and Y.

Regression Coefficients :

The slope of the regression line (8.8.9) is $\rho\frac{\sigma_y}{\sigma_x}$ which is denoted by b_{yx} and it is called the *regression coefficient* of Y on X. Similarly $\rho\frac{\sigma_x}{\sigma_y}$, denoted by b_{xy}, is called the regression coefficient of X on Y.

We find that

$$\rho^2 = b_{xy}\, b_{yx}.$$

$$\text{i.e.,}\quad |\rho| = +\sqrt{b_{xy}\, b_{yx}} \qquad \cdots \qquad \cdots \qquad (8.8.13)$$

and so we can state that $|\rho|$ is equal to the geometric mean of the regression coefficients.

Significance of the correlation coefficient between two random variables X and Y.

Let ρ be the correlation coefficient between X and Y. In (8.8.11) and (8.8.12) we noted that $|\rho|$ is a direct measure of goodness of fit of the regression lines to the joint distribution of X and Y, that is, $|\rho|$ gives a direct measure of concentration of the probability mass near the regression lines. So we can say that the tendency of having a linear relation $Y = c_1^* + c_2^* X$ or $X = d_1^* + d_2^* Y$ increases as $|\rho|$ increases.

Now $0 \leq |\rho| \leq 1$.

So if $|\rho|$ is maximum, that is, $\rho = \pm 1$, the above tendency is maximum and in this case the total probability mass is concentrated on the regression lines which coincide. Again the minimum possible value of $|\rho|$ is 0 and so if $\rho = 0$, the above tendency is minimum, that is, if $\rho = 0$, we get the least possible concentration of the probability mass near the regression lines. If $\rho > 0$, then the equations (8.8.9) and (8.8.10) indicate that it is most likely to get the relation such as 'Y increases as X increases' and if $\rho < 0$, the same equations indicate that it is most likely to have the relation 'Y decreases as X increases'.

We have $\rho = \dfrac{\text{cov}(X, Y)}{\sigma_x\, \sigma_y}$. Then for given σ_x, σ_y we see that $\text{cov}(X, Y)$ and ρ both measure the above mentioned tendency of having linear relationship between X and Y and we note that ρ is a dimensionless measure of the same tendency.

The following figures illustrate the regression lines for different values of ρ.

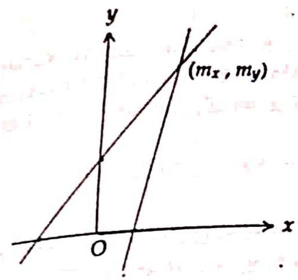


Fig. 8.8.1 ($\rho > 0$)

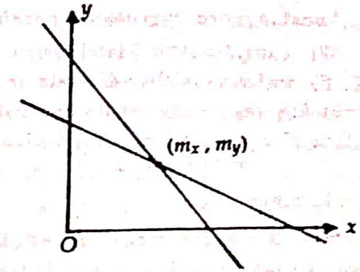


Fig. 8.8.2 ($\rho < 0$)

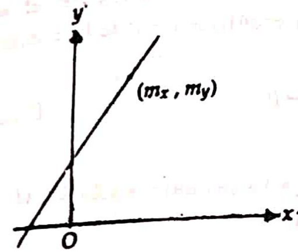


Fig. 8.8.3

($\rho = 1$, when the two regression lines coincide with the straight line passing through (m_x, m_y) and having a slope $\dfrac{\sigma_y}{\sigma_x} > 0$)

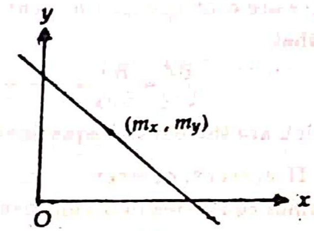


Fig. 8.8.4

($\rho = -1$, when the two regression lines coincide with the straight line passing through (m_x, m_y) and having a slope $-\dfrac{\sigma_y}{\sigma_x} < 0$.)

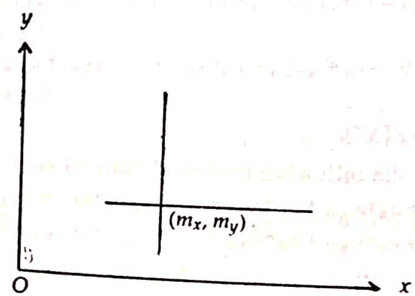


Fig. 8.8.5 ($\rho = 0$)

Least square regression parabola :

We consider the distribution of the two-dimensional variate (X, Y) which is either discrete or continuous. To find the equation of the kth degree regression parabola of Y on X, we are to find the values of $c_1, c_2, \ldots, c_{k+1}$ for which

$$E\left[\{Y-(c_1+c_2X+c_3X^2+\cdots+c_{k+1}X^k)\}^2\right]$$

is minimum.

Let $A=E\left[\{Y-(c_1+c_2X+c_3X^2+\cdots\cdots+c_{k+1}X^k)\}^2\right]$, where we assume that $E(Y), E(Y^2), E(X^r)$ for $r=1, 2, \cdots, 2k$ and $E(X^rY)$ exist for $r=1, 2, \cdots, k$. Then A can be regarded as a function of $k+1$ real variables $c_1, c_2, \cdots, c_{k+1}$ where A and its partial derivatives of any order are continuous. Now a necessary condition for A to be minimum is that

$$\frac{\partial A}{\partial c_1}=\frac{\partial A}{\partial c_2}=\cdots\cdots=\frac{\partial A}{\partial c_{k+1}}=0, \tag{8.8.14}$$

which are the normal equations.

If $c_1=c_1^*, c_2=c_2^*, \ldots, c_{k+1}=c_{k+1}^*$ be the unique solution of the normal equations (8.8.14), then we have

$$\left[\frac{\partial A}{\partial c_i}\right]_{(c_1^*, c_2^*, \ldots, c_{k+1}^*)}=0 \text{ for } i=1, 2, \cdots, k+1. \tag{8.8.15}$$

Now $\dfrac{\partial A}{\partial c_i}=\dfrac{\partial}{\partial c_i}\left[E\{Y-(c_1+c_2X+\cdots+c_{k+1}X^k)\}^2\right]$

$$=E\left[-2X^{i-1}\{Y-(c_1+c_2X+\cdots+c_{k+1}X^k)\}\right].$$

Then from (8.8.15) we get

$$E\{X^{i-1}(Y-c_1^*-c_2^*X-\cdots-c_{k+1}^*X^k)\}=0$$

or, $E(X^{i-1}Y)-c_1^*E(X^{i-1})-c_2^*E(X^i)-\cdots-c_{k+1}^*E(X^{i+k-1})=0,$ for $i=1, 2, \cdots, k+1$

or, $E(X^{i-1}Y)-c_1^*\alpha_{i-1\,0}-c_2^*\alpha_{i0}-\cdots-c_{k+1}^*\alpha_{i+k-1\,0}=0,$ for $i=1, 2, \cdots, k+1,$

where $\alpha_{r0}=E(X^r)$.

Then we get the following system of normal equations :

$$\begin{aligned}
c_1^*\alpha_{00}+c_2^*\alpha_{10}+c_3^*\alpha_{20}+\cdots+c_{k+1}^*\alpha_{k0}&=\alpha_{01}\\
c_1^*\alpha_{10}+c_2^*\alpha_{20}+c_3^*\alpha_{30}+\cdots+c_{k+1}^*\alpha_{k+1\,0}&=\alpha_{11}\\
\cdots\qquad\cdots\qquad&\cdots\\
\cdots\qquad\cdots\qquad&\cdots\\
c_1^*\alpha_{k0}+c_2^*\alpha_{k+1\,0}+c_3^*\alpha_{k+2\,0}+\cdots+c_{k+1}^*\alpha_{2k\,0}&=\alpha_{k1}.
\end{aligned} \tag{8.8.16}$$

Assuming that the system of equations (8.8.16) has a unique solution, solving the system (8.8.16), we can find $c_1^*, c_2^*, \ldots, c_{k+1}^*$ and it can be shown that A is minimum for $c_1=c_1^*, c_2=c_2^*, \ldots, c_{k+1}=c_{k+1}^*$. Then the equation of the kth degree regression parabola of Y on X is $y=c_1^*+c_2^*x+c_3^*x^2+\cdots+c_{k+1}^*x^k$, which is the best fitting parabola of degree k, to the distribution of (X, Y) from the family of parabolas given by $y=c_1+c_2x+c_3x^2+\cdots+c_{k+1}x^k$, where $c_1, c_2, \cdots, c_{k+1}$ are parameters. Similarly the equation of the kth degree regression parabola of X on Y can be found.

Now let A^* be the value of A, where $c_1=c_1^*, c_2=c_2^*, \cdots, c_{k+1}=c_{k+1}^*$. Let

$$B=E\{(tY-c_1-c_2X-c_3X^2-\cdots-c_{k+1}X^k)^2\}.$$

Then B can be regarded as a homogeneous function of $t, c_1, c_2, \ldots, c_{k+1}$ of degree 2. Then by Euler's theorem, we get

$$t\frac{\partial B}{\partial t}+c_1\frac{\partial B}{\partial c_1}+\cdots+c_{k+1}\frac{\partial B}{\partial c_{k+1}}=2B.$$

Now we see that

$$B=A^* \text{ for } t=1, c_1=c_1^*, c_2=c_2^*, \cdots, c_{k+1}=c_{k+1}^*.$$

$$\therefore\ 2A^*=\left[t\frac{\partial B}{\partial t}+c_1\frac{\partial B}{\partial c_1}+c_2\frac{\partial B}{\partial c_2}+\cdots+c_{k+1}\frac{\partial B}{\partial c_{k+1}}\right]_{(t=1, c_1=c_1^*, \ldots, c_{k+1}=c_{k+1}^*)}$$

$$=2E\{Y(Y-c_1^*-c_2^*X-c_3^*X^2-\cdots-c_{k+1}^*X^k)\},$$

since $\dfrac{\partial B}{\partial c_1}=\dfrac{\partial B}{\partial c_2}=\cdots=\dfrac{\partial B}{\partial c_{k+1}}=0$, for $t=1, c_1=c_1^*, c_2=c_2^*, \cdots, c_{k+1}=c_{k+1}^*$, by virtue of normal equations.

So $A^*=E(Y^2)-c_1^*E(Y)-c_2^*E(XY)-\cdots-c_{k+1}^*E(X^kY)$,

i.e., $A^*=\alpha_{02}-c_1^*\alpha_{01}-c_2^*\alpha_{11}-\cdots-c_{k+1}^*\alpha_{k1}$. (8.8.17)

Then A^*, given by (8.8.17), can be taken as an inverse measure of goodness of fit of the regression parabola $y=c_1^*+c_2^*x+\cdots+c_{k+1}^*x^k$ to the distribution of (X, Y).

Now to obtain a direct measure of goodness of fit we proceed as follows :

Let $U=c_1^*+c_2^*X+\cdots+c_{k+1}^*X^k$.

Case I. $c_1^*, c_2^*, \ldots, c_{k+1}^*$ are not all zero.

Then $A^* = E\{(Y-U)^2\}$. Let u be the real variable corresponding to U. Then considering the family of straight lines $y=c+du$, in the (u, y) plane (c, d are parameters), we see that the equation of the regression line of Y on U can be obtained by minimizing

$$E\{(Y-c-dU)^2\} = E\{(Y-c-dc_1^* - dc_2^*X - \cdots - dc_{k+1}^*X^k)^2\}.$$

Now since $E\{(Y-(c_1+c_2X+\cdots+c_{k+1}X^k))^2\}$ is minimum for $c_1=c_1^*, c_2=c_2^*, \ldots, c_{k+1}=c_{k+1}^*$, so

$$E\{(Y-c-dc_1^* - dc_2^*X - \cdots - dc_{k+1}^*X^k)^2\}$$

is minimum for $c=0$, $d=1$.

Hence, the equation of the regression line of Y on U is $y=u$ and by (8.8.11) we get $A^*=\sigma_y^2[1-\{\rho(U, Y)\}^2]$. (8.8.18)

Now the equation of the regression line of Y on U is

$$y-m_y=\rho(U, Y)\cdot\frac{\sigma_y}{\sigma_u}(u-m_u),$$

where $m_u=E(U)$, σ_u is the standard deviation of U.

But here the regression line is $y=u$. Hence we have

$$\rho(U, Y)\frac{\sigma_y}{\sigma_u}=1 \quad \cdots \quad (8.8.19)$$

and

$$m_y=\rho(U, Y)\frac{\sigma_y}{\sigma_u}m_u \quad \cdots \quad (8.8.20)$$

Since $\sigma_y>0$, $\sigma_u>0$, (8.8.19) shows that $\rho(U, Y)>0$.

Case II. $c_1^*=0$, $c_2^*=0$, ..., $c_{k+1}^*=0$.

In this case $U=0$ and $Y=0$ is the best fitting curve to the joint distribution of U and Y from the given family of curves. Now $y=0$ is a straight line and hence in this case $y=0$ is the regression line of Y on U (*i.e.*, the best fitting line to the aforesaid distribution from the family of straight lines given by $y=c+du$) and so in this case by (8.8.8.a) and (8.8.8.b), $c^*=m_y-\rho(U, Y)\cdot\frac{\sigma_y}{\sigma_u}m_u=0$ and $d^*=\rho(U, Y)\frac{\sigma_y}{\sigma_u}=0$, which gives $\rho(U, Y)=0$.

So in any case we get

$0\leqslant\rho(U, Y)\leqslant 1$ (since $-1\leqslant\rho(U, Y)\leqslant 1$ in general).

Now from (8.8.18) we get $\{\rho(U, Y)\}^2=1-\frac{A^*}{\sigma_y^2}$,

which shows that $|\rho(U, Y)|=\rho(U, Y)$ increases as A^* decreases and consequently the non-negative number $\rho(U, Y)=\sqrt{1-\frac{A^*}{\sigma_y^2}}$ can be taken as a direct measure of goodness of fit of the regression parabola

$$y=c_1^*+c_2^*x+\cdots+c_{k+1}^*x^k$$

to the joint distribution of X and Y, where A^* is given by (8.8.17).

8.9. Correlation Ratio.

We know that for a bivariate distribution of a continuous variable (X, Y), the regression curve of Y on X is $y=m_Y(x)$, where $m_Y(x)=E(Y|X=x)$. The random variable corresponding to $m_Y(x)$ is $m_Y(X)$. The correlation coefficient $\rho\{m_Y(X), Y\}$ is called the *correlation ratio* of Y on X.

We know that $y=m_Y(x)$ is the best fitting curve to the joint distribution of X and Y from the family of all continuous curves of the form $y=g(x)$. Following the method of obtaining a direct measure of goodness of fit of a kth degree parabola to a bivariate distribution, we can show that the correlation ratio $\rho\{m_Y(X), Y\}$ can be taken as a direct measure of goodness of fit of the regression curve $y=m_Y(x)$ to the joint distribution of (X, Y).

8.10 Illustrative Examples :

Ex. 1. *Two discrete random variables X and Y take the values 1, 2, 3 and the joint probability distribution of X and Y is given by the following table :*

Y \ X	1	2	3
1	0.1	0.1	0.1
2	0.1	0.2	0.1
3	0.1	0.1	0.1

Find expectations and variances of X, Y and $X+Y$.

We have $P(X=1)=0.3$, $P(X=2)=0.4$, $P(X=3)=0.3$.

$$E(X)=1\times0.3+2\times0.4+3\times0.3=2.0.$$

$$E(X^2)=1\times0.3+4\times0.4+9\times0.3=4.6.$$

$$\therefore \text{ var } X=E(X^2)-\{E(X)\}^2$$
$$=4.6-4$$
$$=0.6.$$

Again $P(Y=1)=0.3$, $P(Y=2)=0.4$, $P(Y=3)=0.3$.

$$E(Y)=1\times0.3+2\times0.4+3\times0.3=2.0.$$

$$E(Y^2)=1\times0.3+4\times0.4+9\times0.3=4.6.$$

$$\therefore \text{ var } Y=E(Y^2)-\{E(Y)\}^2=0.6.$$

Let $Z=X+Y$, then Z assumes the values 2, 3, 4, 5, 6.

$$P(Z=2)=0.1,\ P(Z=3)=0.2,\ P(Z=4)=0.4,\ P(Z=5)=0.2$$
$$P(Z=6)=0.1.$$

$$E(X+Y)=2\times0.1+3\times0.2+4\times0.4+5\times0.2+6\times0.1=4.0.$$

$$\begin{aligned}E(XY)&=1\times1\times0.1+1\times2\times0.1+1\times3\times0.1+2\times1\times0.1\\&\quad+2\times2\times0.2+2\times3\times0.1+3\times1\times0.1+3\times2\times0.1\\&\quad+3\times3\times0.1\\&=4.0.\end{aligned}$$

$$\begin{aligned}\text{cov}(X,Y)&=E(XY)-E(X)\,E(Y)\\&=4.0-2\times2\\&=0.\end{aligned}$$

$$\begin{aligned}\therefore \text{ var }(X+Y)&=\text{var } X+\text{var } Y+2\text{ cov }(X,Y)\\&=0.6+0.6\\&=1.2.\end{aligned}$$

Ex. 2. *Two points are dropped at random on a line-segment $(0, a)$. Find the expectation of the square of the distance between the points.*

Let the random variable X, Y denote the numbers giving the position of the two points chosen at random on the given line segment. The density functions $f_X(x)$ and $f_Y(y)$ of X and Y are then given by

$$f_X(x)=\frac{1}{a},\ 0<x<a$$

$$f_Y(y)=\frac{1}{a},\ 0<y<a.$$

X, Y being independent, the joint density function of X and Y is given by

$$f(x,y)=\frac{1}{a^2},\ 0<x<a,\ 0<y<a.$$

Now square of the distance between the two points is

$$Z=(X-Y)^2.$$

$$\therefore\ E(Z)=\int_0^a\int_0^a (x-y)^2\cdot\frac{1}{a^2}\,dx\,dy$$

$$=\frac{1}{a^2}\int_0^a\left[\frac{a^3}{3}-a^2y+ay^2\right]dy=\frac{a^2}{6}.$$

Ex. 3. *Let X, Y be independent normal variates with the common mean m and common variance 1. If*

$$P(X+2Y\leqslant 3)=P(2X-Y\geqslant 4),$$

then find the value of m.

Let $Z=\dfrac{X+2Y}{3}$ and $W=\dfrac{2X-Y}{4}$.

Then $E(Z)=\frac{1}{3}\{E(X)+2E(Y)\}=m$, since $E(X)=E(Y)=m$.

Var $Z=\frac{1}{9}(\text{var } X+4\text{ var } Y)=\frac{5}{9}$, since var $X=$ var $Y=1$ and X, Y are independent.

Again $E(W)=\frac{1}{4}\{2E(X)-E(Y)\}=\dfrac{m}{4}$,

$$\text{var } W=\tfrac{1}{16}(4\text{ var } X+\text{var } Y)=\tfrac{5}{16}.$$

Thus $\dfrac{Z-m}{\frac{\sqrt5}{3}}$ and $\dfrac{W-\frac{m}{4}}{\frac{\sqrt5}{4}}$ are both standard normal variates.

Now the given relation implies

$$P\left(\frac{Z-m}{\frac{\sqrt5}{3}}\leqslant\frac{1-m}{\frac{\sqrt5}{3}}\right)=P\left(\frac{W-\frac{m}{4}}{\frac{\sqrt5}{4}}\geqslant\frac{1-\frac{m}{4}}{\frac{\sqrt5}{4}}\right)$$

$$\text{or, }\int_{-\infty}^{\frac{1-m}{\frac{\sqrt5}{3}}}e^{-\frac{t^2}{2}}\,dt=\int_{\frac{1-\frac{m}{4}}{\frac{\sqrt5}{4}}}^{\infty}e^{-\frac{t^2}{2}}\,dt=\int_{-\infty}^{-\frac{1-\frac{m}{4}}{\frac{\sqrt5}{4}}}e^{-\frac{t^2}{2}}\,dt.$$

$$\therefore\ \frac{1-m}{\frac{\sqrt{5}}{3}}=\frac{\frac{m}{4}-1}{\frac{\sqrt{5}}{4}} \quad \text{or,}\quad 3-3m=m-4.$$

$$\therefore\ m=\tfrac{7}{4}$$

Ex. 4. *Two random variables X, Y have the least square regression lines with equations $3x+2y-26=0$ and $6x+y-31=0$. Find $E(X)$, $E(Y)$ and $\rho(X, Y)$.*

The two lines intersect at $(4, 7)$, which gives

$$E(X)=m_x=4,\ E(Y)=m_y=7.$$

If the regression line of Y on X is

$$3x+2y-26=0,\quad \text{i.e., } y=-\tfrac{3}{2}(x-\tfrac{26}{3}),$$

$$\text{then}\quad \rho\frac{\sigma_y}{\sigma_x}=-\frac{3}{2}$$

Then the regression line of X on Y is

$$6x+y-31=0,\quad \text{i.e., } x=-\tfrac{1}{6}(y-31),$$

so that
$$\rho\frac{\sigma_x}{\sigma_y}=-\tfrac{1}{6}. \tag{8.10.1}$$

$$\therefore\ \left(\rho\frac{\sigma_y}{\sigma_x}\right)\left(\rho\frac{\sigma_x}{\sigma_y}\right)=\tfrac{3}{2}\times\tfrac{1}{6}=\tfrac{1}{4}.$$

$$\text{i.e., } \rho^2=\tfrac{1}{4} \quad \therefore\ \rho=\pm\tfrac{1}{2}.$$

But $\rho=\frac{1}{2}$ does not satisfy (8.10.1), since $\sigma_x>0$, $\sigma_y>0$.

$$\therefore\ \rho=\rho(X, Y)=-\tfrac{1}{2}.$$

Ex. 5. *The probability density functions of two independent random variables X and Y are defined by*

$$f_1(x)=4ax,\quad 0<x<r$$
$$=0\ ,\quad \textit{elsewhere};$$
$$f_2(y)=4by,\quad 0<y<s$$
$$=0\ ,\quad \textit{elsewhere}.$$

Find the value of $\rho(X+Y, X-Y)$.

$$E(X)=\int_0^r x f_1(x)\,dx=\frac{4ar^3}{3}=m_x.$$

$$E(Y)=\frac{4bs^3}{3}=m_y.$$

Let $U=X+Y$, $V=X-Y$. Then

$$\begin{aligned}\operatorname{cov}(U, V)&=E\left[\{(X+Y)-(m_x+m_y)\}\{(X-Y)-(m_x-m_y)\}\right]\\&=E\left[\{(X-m_x)+(Y-m_y)\}\{(X-m_x)-(Y-m_y)\}\right]\\&=E\{(X-m_x)^2\}-E\{(Y-m_y)^2\}\\&=\operatorname{var}X-\operatorname{var}Y.\end{aligned}$$

$$E(X^2)=\int_0^r x^2 f_1(x)\,dx=ar^4,$$

$$E(Y^2)=bs^4.$$

$$\therefore\ \operatorname{var}X=E(X^2)-\{E(X)\}^2=ar^4-\frac{16a^2r^6}{9},$$

$$\operatorname{var}Y=bs^4-\frac{16b^2s^6}{9},$$

$$\therefore\ \operatorname{cov}(U, V)=ar^4-bs^4-\tfrac{16}{9}(a^2r^6-b^2s^6).$$

Again $\operatorname{var}U=\operatorname{var}X+\operatorname{var}Y=\operatorname{var}V$.

$$\therefore\ \operatorname{var}U=ar^4+bs^4-\tfrac{16}{9}(a^2r^6+b^2s^6)=\operatorname{var}V.$$

$$\therefore\ \rho(U, V)=\frac{\operatorname{cov}(U, V)}{\sqrt{\operatorname{var}U}\sqrt{\operatorname{var}V}}$$

$$=\frac{ar^4-bs^4-\frac{16}{9}(a^2r^6-b^2s^6)}{ar^4+bs^4-\frac{16}{9}(a^2r^6+b^2s^6)}.$$

Now $\int_{-\infty}^{\infty} f_1(x)\,dx=1$ gives $\int_0^r 4ax\,dx=1$, i.e., $r^2=\dfrac{1}{2a}$,

and $\int_{-\infty}^{\infty} f_2(y)\,dy=1$ gives $s^2=\dfrac{1}{2b}$.

$$\text{Hence, } \rho(U, V)=\frac{\frac{1}{4a}-\frac{1}{4b}-\frac{2}{9a}+\frac{2}{9b}}{\frac{1}{4a}+\frac{1}{4b}-\frac{2}{9a}-\frac{2}{9b}}=\frac{\frac{1}{36a}-\frac{1}{36b}}{\frac{1}{36a}+\frac{1}{36b}}$$

$$=\frac{b-a}{b+a}.$$

Ex. 6. *Calculate the covariance between X and Y, if the joint probability density function is given by*

$$f(x, y)=e^{-x-y},\quad x>0,\ y>0$$
$$=0\ ,\quad \textit{elsewhere}.$$

[C. H. (Econ.) '82]

Here $m_x = E(X) = \int_0^\infty \int_0^\infty x e^{-x-y}\,dx\,dy$

$$= \int_0^\infty \left(x e^{-x} \int_0^\infty e^{-y}\,dy \right) dx$$

$$= \int_0^\infty x e^{-x}\,dx$$

$$= \Gamma(2) = 1.$$

Similarly, $m_y = E(Y) = 1.$

$$E(XY) = \int_0^\infty \int_0^\infty xy\, e^{-x-y}\,dx\,dy$$

$$= \int_0^\infty \left(x e^{-x} \int_0^\infty y e^{-y}\,dy \right) dx$$

$$= \int_0^\infty x e^{-x}\,dx = 1.$$

So, the required covariance is given by

$$\text{cov}(X, Y) = E(XY) - m_x m_y = 1 - 1 = 0.$$

Ex. 7. *If σ_x^2, σ_y^2, σ_{x-y}^2 are the variances of X, Y and $X-Y$ respectively and r is the correlation coefficient of X and Y, then prove that*

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}.$$ [C. H. (Math) '89]

If m_x, m_y be the respective means of the random variables X and Y, we get

$$E(X-Y) = E(X) - E(Y) = m_x - m_y,$$

mean of the random variable $X-Y$

$$\therefore\ E\{(X-Y)-(m_x-m_y)\}^2 = E\{(X-m_x)-(Y-m_y)\}^2$$

$$= E\{(X-m_x)^2\} + E\{(Y-m_y)^2\} - 2E\{(X-m_x)(Y-m_y)\}$$

or, $\text{var}(X-Y) = \text{var}\,X + \text{var}\,Y - 2\,\text{cov}(X, Y)$

or, $\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$.

$$\therefore\ r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}.$$

Ex. 8. *The probability density function of (X, Y) is*

$$f(x, y) = \frac{1}{\pi\sqrt{3}}\, exp \left\{ -2\left(x^2 - xy + \frac{y^2}{3} \right) \right\},$$

$$-\infty < x < \infty,\quad -\infty < y < \infty.$$

Find the regression curve of Y on X. [C. H. (Math.) '95]

The regression curve of Y on X is $y = m_Y(x)$, where $m_Y(x)$ is the conditional mean of Y on the hypothesis $X = x$ and it is given by

$$m_Y(x) = \frac{\int_{-\infty}^{\infty} y f(x, y)\,dy}{f_X(x)}.$$

Now $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,dy$

$$= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} e^{-2\left(x^2 - xy + \frac{y^2}{3}\right)}\,dy$$

$$= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} e^{-2\left\{\left(\frac{y}{\sqrt{3}} - \frac{\sqrt{3}}{2}x\right)^2 + \frac{x^2}{4}\right\}}\,dy$$

$$= \frac{e^{-\frac{x^2}{2}}}{\pi\sqrt{3}} \int_{-\infty}^{\infty} e^{-2\left(\frac{y}{\sqrt{3}} - \frac{\sqrt{3}}{2}x\right)^2}\,dy$$

$$= \frac{e^{-\frac{x^2}{2}}}{\pi\sqrt{2}} \int_{-\infty}^{\infty} e^{-z^2}\,dz$$

where $z = \sqrt{2}\left(\frac{y}{\sqrt{3}} - \frac{\sqrt{3}}{2}x \right)$

$$= \frac{e^{-\frac{x^2}{2}}}{\pi\sqrt{2}} \times \sqrt{\pi} = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Again $\int_{-\infty}^{\infty} y f(x, y)\,dy$

$$= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} y\, e^{-2\left(x^2 - xy + \frac{y^2}{3}\right)}\,dy$$

$$= \frac{1}{\pi\sqrt{3}} \int_{-\infty}^{\infty} y\, e^{-2\left\{\left(\frac{y}{\sqrt{3}} - \frac{\sqrt{3}}{2}x\right)^2 + \frac{x^2}{4}\right\}}\,dy$$

or. $$\int_{-\infty}^{\infty} y\, f(x, y)\, dy = \frac{e^{-\frac{x^2}{2}}}{\pi\sqrt{3}} \int_{-\infty}^{\infty} y\, e^{-2\left(\frac{y}{\sqrt{3}} - \frac{\sqrt{3}}{2} x\right)^2} dy$$

$$= \frac{e^{-\frac{x^2}{2}}}{\pi\sqrt{2}} \sqrt{3} \int_{-\infty}^{\infty} \left(\frac{z}{\sqrt{2}} + \frac{\sqrt{3}}{2} x\right) e^{-z^2}\, dz,$$

where $z = \sqrt{2}\left(\frac{y}{\sqrt{3}} - \frac{\sqrt{3}}{2} x\right)$

$$= \frac{e^{-\frac{x^2}{2}}}{\pi\sqrt{2}} \sqrt{3} \left(\frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} z\, e^{-z^2}\, dz + \frac{\sqrt{3}}{2} x \int_{-\infty}^{\infty} e^{-z^2}\, dz \right).$$

Now $\int_{-\infty}^{\infty} z\, e^{-z^2}\, dz$ is convergent and its value is zero.

$$\therefore \int_{-\infty}^{\infty} y\, f(x, y)\, dy$$

$$= \frac{\sqrt{3}\, e^{-\frac{x^2}{2}}}{\pi\sqrt{2}} \cdot \frac{\sqrt{3}}{2} x \int_{-\infty}^{\infty} e^{-z^2}\, dz$$

$$= \frac{3x\, e^{-\frac{x^2}{2}}}{2\sqrt{2}\,\pi} \sqrt{\pi} = \frac{3x\, e^{-\frac{x^2}{2}}}{2\sqrt{2}\sqrt{\pi}}.$$

$$\therefore\ m_Y(x) = \frac{\int_{\infty}^{\infty} y\, f(x, y)\, dy}{f_X(x)} = \frac{\dfrac{3x\, e^{-\frac{x^2}{4}}}{\sqrt{2}\sqrt{\pi}}}{\dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}} = \frac{3}{2} x.$$

$\therefore$ the regression curve of Y on X is $y = \frac{3}{2} x$.

Ex. 9. *For the continuous distribution defined by*

$$f(x, y) = 3x^2 - 8xy + 6y^2,\ 0 < x < 1,\ 0 < y < 1,$$

find the regression curves for the means and also the least square regression lines. [*C. H. (Math.)* '87]

The marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \int_0^1 (3x^2 - 8xy + 6y^2)\, dy$$

$$= 3x^2 - 4x + 2,\ 0 < x < 1;$$

and that of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

$$= \int_0^1 (3x^2 - 8xy + 6y^2)\, dx$$

$$= 6y^2 - 4y + 1,\ 0 < y < 1.$$

Again $$\int_{-\infty}^{\infty} y\, f(x, y)\, dy$$

$$= \int_0^1 y(3x^2 - 8xy + 6y^2)\, dy = \tfrac{1}{6}(9x^2 - 16x + 9)$$

and $$\int_{-\infty}^{\infty} x\, f(x, y)\, dx$$

$$= \int_0^1 x(3x^2 - 8xy + 6y^2)\, dx = \tfrac{1}{12}(36y^2 - 32y + 9).$$

$$\therefore\ m_X(x) = \frac{\int_{-\infty}^{\infty} y\, f(x, y)\, dy}{f_X(x)} = \frac{9x^2 - 16x + 9}{6(3x^2 - 4x + 2)}$$

and $$m_Y(y) = \frac{\int_{-\infty}^{\infty} x\, f(x, y)\, dx}{f_Y(y)} = \frac{36y^2 - 32y + 9}{12(6y^2 - 4y + 1)}.$$

Hence the regression curve for the mean of Y is

$$6y(3x^2 - 4x + 2) = 9x^2 - 16x + 9$$

and that for the mean of X is

$$12x(6y^2 - 4y + 1) = 36y^2 - 32y + 9.$$

Again the means m_x and m_y for the marginal distributions of X and Y are respectively

$$m_x = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_0^1 x(3x^2 - 4x + 2)\, dx = \tfrac{5}{12}$$

and $$m_y = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_0^1 y(6y^2 - 4y + 1)\, dy = \tfrac{2}{3}$$

The variances σ_x^2 and σ_y^2 are given by

$$\sigma_x^2 = E(X^2) - m_x^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - \frac{25}{144}$$

$$= \int_0^1 x^2(3x^2 - 4x + 2)\, dx - \frac{25}{144}$$

$$= \frac{4}{15} - \frac{25}{144} = \frac{67}{720}$$

and $\quad \sigma_y^2 = E(Y^2) - m_y^2 = \int_0^1 y^2(6y^2 - 4y + 1)\,dy - \frac{4}{9} = \frac{4}{45}.$

$$E(XY) = \int_0^1 \int_0^1 xy(3x^2 - 8xy + 6y^2)\,dx\,dy$$

$$= \int_0^1 \left\{ y \int_0^1 (3x^3 - 8x^2 y + 6y^2 x)\,dx \right\} dy$$

$$= \frac{1}{12} \int_0^1 y(36y^2 - 32y + 9)\,dy = \frac{17}{72}.$$

$\therefore \quad \operatorname{cov}(X, Y) = E(XY) - m_x m_y$

$$= \frac{17}{72} - \left(\frac{5}{12} \times \frac{2}{3}\right) = -\frac{1}{24}.$$

$\therefore \quad b_{yx} = \rho \dfrac{\sigma_y}{\sigma_x} = \dfrac{\operatorname{cov}(X, Y)}{\sigma_x^2} = -\dfrac{30}{37}$

and $\quad b_{xy} = \rho \dfrac{\sigma_x}{\sigma_y} = \dfrac{\operatorname{cov}(X, Y)}{\sigma_y^2} = -\dfrac{15}{32}.$

Hence the regression lines are

$$y - \frac{2}{3} = -\frac{30}{37}\left(x - \frac{5}{12}\right)$$

and

$$x - \frac{5}{12} = -\frac{15}{32}\left(y - \frac{2}{3}\right).$$

Ex. 10. *X, Y and Z are three random variables so that X and Y are independent and $Z = XY$. X can take two values 10 and 20 and the probability that X is 10 is $\frac{1}{3}$. Y can take three values 5, 6, 7. The probability that Y is 5 is $\frac{1}{4}$ and that it is 6 is $\frac{1}{2}$. Find the expectation of Z.* [C. H. (Econ.) '88]

$$P(X = 10) = \tfrac{1}{3}, \; P(X = 20) = 1 - \tfrac{1}{3} = \tfrac{2}{3}.$$

$\therefore \quad E(X) = 10 \times \frac{1}{3} + 20 \times \frac{2}{3} = \frac{50}{3}.$

$$P(Y = 5) = \tfrac{1}{4}, \; P(Y = 6) = \tfrac{1}{2}, \; P(Y = 7) = 1 - (\tfrac{1}{4} + \tfrac{1}{2}) = \tfrac{1}{4}.$$

$\therefore \quad E(Y) = 5 \times \frac{1}{4} + 6 \times \frac{1}{2} + 7 \times \frac{1}{4} = 6.$

Since $Z = XY$, where X and Y are independent,

$$E(Z) = E(XY) = E(X)E(Y) = \tfrac{50}{3} \times 6 = 100.$$

Ex. 11. *If X, Y are independent standard normal variates, then*

$$E\{\min(|X|, |Y|)\} = 4\int_0^\infty \{1 - \Phi(x)\}^2\,dx,$$

where $\Phi(x)$ is the distribution function of a standard normal variate.

$E\{\min(|X|, |Y|)\}$

$= \iint_{|x| > |y|} |y|\, f(x, y)\,dx\,dy + \iint_{|y| > |x|} |x|\, f(x, y)\,dx\,dy$

$= 2\iint_{|y| > |x|} |x|\, f(x, y)\,dx\,dy$, from symmetry, where $f(x, y)$ is the joint density function of the two-dimensional random variable (X, Y). Now X and Y being independent,

$$f(x, y) = f_X(x)\, f_Y(y).$$

$f_X(x)$, $f_Y(y)$ are respectively the marginal density functions of X and Y.

$$\therefore \quad f(x, y) = \Phi'(x)\, \Phi'(y),$$

X and Y being both standard normal variates and $\Phi(x)$ is the distribution function of a standard normal variate.

$\therefore \quad E\{\min(|X|, |Y|)\}$

$$= 2\iint_{|y| > |x|} |x|\, \Phi'(x)\, \Phi'(y)\,dx\,dy$$

$$= 4\int_{-\infty}^{\infty} |x|\, \Phi'(x) \left\{ \int_{|x|}^{\infty} \Phi'(y)\,dy \right\} dx$$

$$= 8\int_0^\infty x\, \Phi'(x) \left\{ \int_x^\infty \Phi'(y)\,dy \right\} dx,$$

from symmetry of a normal distribution

$$= 8\int_0^\infty x\, \Phi'(x) \{1 - \Phi(x)\}\,dx$$

$$= 8 \operatorname*{Lt}_{B \to \infty} \left[-\tfrac{1}{2} x \{1 - \Phi(x)\}^2 \right]_0^B + 8 \cdot \tfrac{1}{2} \int_0^\infty 1 \cdot \{1 - \Phi(x)\}^2\,dx$$

(integrating by parts)

$$= 4\int_0^\infty \{1 - \Phi(x)\}^2\,dx.$$

Ex. 12. *Find the mathematical expectation of the total number of points in a bridge hand of 13 cards where the points are assigned as follows:*

2 for spade, 4 for heart, 3 for club and 6 for diamond.

Let the random variable X_i denote the point received by the player in the ith drawing $(i = 1, 2, \ldots, 13)$. From symmetry, X_i takes values 2, 4, 3 and 6 each with probability $\frac{1}{4}$.

$\therefore \quad E(X_i) = 2 \times \frac{1}{4} + 4 \times \frac{1}{4} + 3 \times \frac{1}{4} + 6 \times \frac{1}{4} = \frac{15}{4}.$

If S be the random variable, denoting the total number of points, then

$$S = X_1 + X_2 + \cdots\cdots + X_{13}.$$

$$\therefore \quad E(S) = E(X_1) + E(X_2) + \cdots\cdots + E(X_{13})$$
$$= 13 \times \tfrac{15}{4} = \tfrac{195}{4}.$$

Ex. 13. *An urn contains* 100 *tickets numbered* 1, 2, ..., 100, *from which* 10 *tickets are drawn successively without replacement. Find the mean and variance of the sum of the numbers on the tickets drawn-*

Let the random variable X_i denote the number of the ith ticket drawn, $i = 1, 2, \ldots, 10$. From symmetry, X_i can take any one of the numbers 1, 2, ..., 100 and the probability of its taking any one of the 100 numbers is equal to $\frac{1}{100}$. So

$$E(X_i) = \frac{1 + 2 + \cdots + 100}{100} = \frac{101}{2}.$$

Let $S = X_1 + X_2 + \cdots + X_{10}$ be the random variable denoting the sum of the numbers on the tickets drawn.

$$\therefore \quad E(S) = E(X_1) + E(X_2) + \cdots\cdots + E(X_{10})$$
$$= \frac{10 \times 101}{2} = 505.$$

Again, $\operatorname{var} X_i = E(X_i^2) - \{E(X_i)\}^2$

$$= \frac{1^2 + 2^2 + \cdots + 100^2}{100} - \left(\frac{101}{2}\right)^2$$
$$= \frac{100 \times 101 \times 201}{600} - \left(\frac{101}{2}\right)^2 = \frac{3333}{4}.$$

We now consider the distribution of the random variable (X_1, X_2). The spectrum is (i, j) for $i = 1, 2, \cdots, 100$; $j = 1, 2, \cdots, 100$. Since the tickets are drawn without replacement

$$P(X_1 = i, X_2 = j) = \frac{1}{100 \times 99} \qquad \text{if } i \neq j$$
$$= 0 \qquad \text{if } i = j$$

$$\therefore \quad E(X_1 X_2) = \frac{1}{100 \times 99}\left(\sum_{j=1}^{100}\sum_{i=1}^{100} ij - \sum_{i=1}^{100} i^2\right)$$
$$= \frac{1}{100 \times 99}\left\{\left(\frac{100 \times 101}{2}\right)^2 - \frac{100 \times 101 \times 201}{6}\right\} = \frac{15251}{6}.$$

$$\therefore \quad \operatorname{cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$$
$$= \frac{15251}{6} - \left(\frac{101}{2}\right)^2 = -\frac{101}{12}.$$

Then from symmetry, $\operatorname{cov}(X_i, X_j) = -\frac{101}{12}$, $i \neq j$;
$i = 1, 2, \cdots, 100$; $j = 1, 2, \cdots, 100$.

$$\text{Hence,} \quad \operatorname{var} S = \sum_{i=1}^{10} \operatorname{var} X_i + 2\sum_{i<j} \operatorname{cov}(X_i, X_j)$$
$$= 10 \times \frac{3333}{4} + 2\left(-\frac{101}{12}\right)(9 + 8 + 7 + \cdots + 1) = 7575.$$

Ex. 14. *If the independent random variables* $X_1, X_2, \ldots, \ldots, X_n$ *all have the same distribution and their sum is normally distributed, then prove that each of them is normally distributed.* [C. H. (Math.) '91]

Let $S_n = X_1 + X_2 + \cdots + X_n$ be normal (m, σ) distributed, where $X_1, X_2, \ldots, X_n$ are independent random variables, each having the same distribution. Let $\chi(t)$ be the characteristic function of each of the random variables $X_i (i = 1, 2, \ldots, n)$ and $K(t)$ be that of S_n. Then

$$K(t) = e^{imt - \frac{1}{2}\sigma^2 t^2}.$$

Also $K(t) = E(e^{itS_n})$

$$\therefore \quad e^{imt - \frac{1}{2}\sigma^2 t^2} = E\{e^{it(X_1 + X_2 + \cdots + X_n)}\}$$
$$= E(e^{itX_1} \cdot e^{itX_2} \cdots\cdots e^{itX_n})$$
$$= E(e^{itX_1})\, E(e^{itX_2}) \cdots\cdots E(e^{itX_n}),$$
since $X_1, X_2, \cdots, X_n$ are independent
$$= \{\chi(t)\}^n.$$

$$\therefore \quad \chi(t) = e^{i\frac{m}{n}t - \frac{1}{2}\left(\frac{\sigma}{\sqrt{n}}\right)^2 t^2}$$

which shows that $X_1, X_2, \ldots, X_n$ are each normally $\left(\frac{m}{n}, \frac{\sigma}{\sqrt{n}}\right)$ distributed.

Ex 15. *The joint probability density function of the random variables* X *and* Y *is*

$$f(x, y) = k(1 - x - y) \text{ for } x \geqslant 0,\ y \geqslant 0,\ x + y < 1$$
$$= 0 \text{ elsewhere,}$$

where k *is a constant.*

Find (*i*) *the mean value of* Y *when* $X = \frac{1}{2}$.

(*ii*) *the covariance of* X *and* Y. [C. H. (*Math.*) '92]

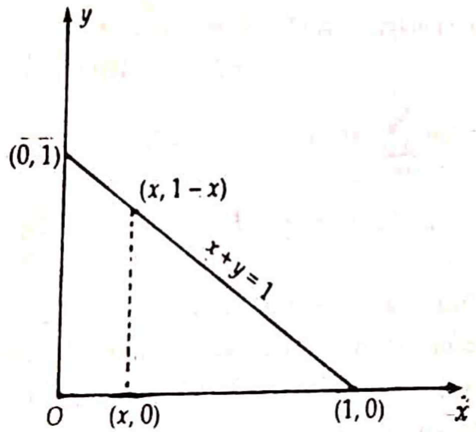


Fig. 8.10.1

From $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$, we get

$$k\int_0^1\left\{\int_0^{1-x}(1-x-y)\,dy\right\}dx = 1$$

or, $\frac{k}{2}\int_0^1 (1-x)^2 dx = 1$ or, $k = 6$.

If $f_X(x)$ and $f_Y(y)$ be the marginal density functions of X and Y, then

$$f_X(x) = 6\int_{-\infty}^{\infty} f(x, y)\,dy = 6\int_0^{1-x}(1-x-y)\,dy = 3(1-x)^2,\ 0<x<1.$$

$$f_Y(y) = 6\int_{-\infty}^{\infty} f(x, y)\,dx = 6\int_0^{1-y}(1-x-y)\,dx = 3(1-y)^2,\ 0<y<1.$$

$$\therefore\ E(X) = \int_{-\infty}^{\infty} x\,f_X(x) = 3\int_0^1 x(1-x)^2 dx = 3B(2, 3) = \frac{3\Gamma(2)\Gamma(3)}{\Gamma(5)} = \frac{1}{4},$$

$$E(Y) = 3\int_0^1 y(1-y)^2 dy = \frac{1}{4}.$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 6xy\,(1-x-y)\,dx\,dy$$

$$= 6\int_0^1\left\{\int_0^{1-x} xy(1-x-y)\,dy\right\}dx$$

$$= \int_0^1 x(1-x)^3\,dx = B(2, 4) = \frac{\Gamma(2)\Gamma(4)}{\Gamma(6)} = \frac{1}{20}.$$

$$\therefore\ \operatorname{cov}(X, Y) = E(XY) - E(X)\,E(Y)$$

$$= \frac{1}{20} - \frac{1}{16} = -\frac{1}{80}.$$

The conditional mean of Y when $X = \frac{1}{2}$ is given by

$$m_Y(\tfrac{1}{2}) = E(Y\,|\,X = \tfrac{1}{2}) = \frac{\int_{-\infty}^{\infty} y\,f(\frac{1}{2}, y)\,dy}{f_X(\frac{1}{2})}$$

$$= \frac{6\int_0^{\frac{1}{2}} y(1-\frac{1}{2}-y)\,dy}{3(1-\frac{1}{2})^2}$$

$$= 4\int_0^{\frac{1}{2}} y(1-2y)\,dy = -\frac{1}{6}$$

Ex. 16. *The random variables* X, Y *are connected by the linear relation* $2X + 3Y + 4 = 0$. *Show that* $\rho(X, Y) = -1$.

Here $2X + 3Y + 4 = 0$. (8.10.2)

$\therefore\ E(2X + 3Y + 4) = 0$ (8.10.3)

or, $2m_x + 3m_y + 4 = 0$, where $m_x = E(X)$, $m_y = E(Y)$.

Then from (8.10.2) and (8.10.3) we get

$$2(X - m_x) + 3(Y - m_y) = 0$$

or, $2(X - m_x) = -3(Y - m_y)$. (8.10.4)

$\therefore\ 4E\{(X - m_x)^2\} = 9E\{(Y - m_y)^2\}$

or, $4\sigma_x^2 = 9\sigma_y^2$

or, $2\sigma_x = 3\sigma_y$, since $\sigma_x > 0$, $\sigma_y > 0$. (8.10.5)

Hence, $\rho(X, Y) = \dfrac{E\{(X - m_x)(Y - m_y)\}}{\sigma_x\,\sigma_y}$

$$= \frac{E\{-\frac{2}{3}(X - m_x)^2\}}{\sigma_x \cdot \frac{2}{3}\sigma_x}, \text{ by (8.10.4) and (8.10.5)}$$

$$= \frac{-\frac{2}{3}\sigma_x^2}{\frac{2}{3}\sigma_x^2} = -1.$$

Ex. 17. *Show that* $2X + 3Y$ *and* $4X + 9Y$ *are uncorrelated if*

$$8\sigma_x^2 + 30\rho\sigma_x\sigma_y + 27\sigma_y^2 = 0.$$

Let $U = 2X + 3Y$, $V = 4X + 9Y$.

Then $m_u = 2m_x + 3m_y$, $\quad m_v = 4m_x + 9m_y$.

$$\therefore \quad (U-m_u)(V-m_v)$$
$$= \{(2X+3Y)-(2m_x+3m_y)\}\times\{(4X+9Y)-(4m_x+9m_y)\}$$
$$= \{2(X-m_x)+3(Y-m_y)\}\{4(X-m_x)+9(Y-m_y)\}$$
$$= 8(X-m_x)^2 + 27(Y-m_y)^2 + 30(X-m_x)(Y-m_y).$$
$$\therefore \quad \text{cov}\,(U, V) = E\{(U-m_u)(V-m_v)\}$$
$$= 8E\{(X-m_x)^2\} + 27E\{(Y-m_y)^2\} + 30\,E\{(X-m_x)(Y-m_y)\}$$
$$= 8\sigma_x^2 + 27\sigma_y^2 + 30\,\text{cov}\,(X, Y)$$
$$= 8\sigma_x^2 + 27\sigma_y^2 + 30\rho\sigma_x\sigma_y.$$

Hence, U, V are uncorrelated if $8\sigma_x^2 + 30\rho\sigma_x\sigma_y + 27\sigma_y^2 = 0$.

Ex. 18. *Two points are independently chosen at random on two sides of a square, the length of a side being b. Find the mean area of the triangle formed by the line joining the two random points and the sides of the square.*

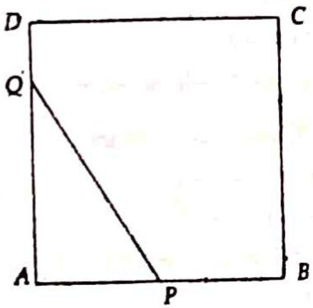


Fig. 8.10 2

Let X and Y be the random variables corresponding to the lengths AP and AQ, where P and Q are the two points chosen at random on the two sides AB and AD of the square $ABCD$ of side b. Then X and Y are independent random variables, each uniformly distributed in $[0, b]$, so that their marginal density functions are given by

$$f_X(x) = \frac{1}{b}, \; 0 < x < b$$
$$= 0, \text{ elsewhere.}$$

and

$$f_Y(y) = \frac{1}{b}, \; 0 < y < b$$
$$= 0, \text{ elsewhere.}$$

Area of the triangle $APB = \frac{1}{2}XY$.

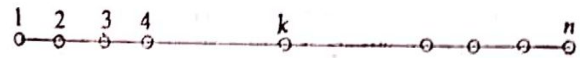
$\therefore$ if $f(x, y)$ be the joint density function of the random variables X and Y, then the required mean area is given by

$$E(\tfrac{1}{2}XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tfrac{1}{2}xy\, f(x, y)\, dx\, dy$$
$$= \tfrac{1}{2}\left\{\int_0^b x f_X(x)\, dx\right\}\left\{\int_0^b y f_Y(y)\, dy\right\},$$

X and Y being independent

$$= \frac{1}{2b^2}\left(\int_0^b x\, dx\right)\left(\int_0^b y\, dy\right) = \frac{b^2}{8}.$$

Ex. 19. *A workman is operating n machines of same type arranged in a straight line at separation a from one another. Assuming that an operator moves from one machine to another machine in order of priority, find the average path length between the machines.*



Let the machines be numbered 1 to n from left to right. Let the random variables X and Y denote respectively the number of the machine presently attended by the workman and the number of the next machine requiring the attention of the workman according to priority. The spectrum of X is

$$x_i = i, \; (i = 1, 2, \cdots, n)$$

and that Y is

$$y_j = j, \; (j = 1, 2, \cdots, n).$$

Since the machines are of the same type,

$$P(X = x_i) = \frac{1}{n} \quad (1 < i < n).$$

We are to find the average path length, *i.e.*, the expectation of the path length denoted by the random variable Z.

Let Z_k^i be the path length between machines $X = x_i$ and $Y = y_k$ (i.e. path length when kth machine is required to be attended next by the operator when his present position is the ith machine);

$$i = 1, 2, \cdots, n\,;\; j = 1, 2, \cdots, n.$$

Then $Z_k^i = \begin{cases} (i-k)a & \text{when } i > k \\ (k-i)a & \text{when } i < k. \end{cases}$

From definition, $E(Z \mid X=x_i) = \sum_{k=1}^{n} Z_k^i\, P(Y=k \mid X=i)$

$$= \sum_{k=1}^{i} \frac{1}{n}(i-k)a + \sum_{k=i+1}^{n} (k-i)a \cdot \frac{1}{n},$$

since $P(Y=k \mid X=i) = \frac{1}{n}$

$$= \frac{1}{n}\,[a\{(i-1)+(i-2)+\cdots\cdots+1\} + a\{1+2+\cdots\cdots+(n-i)\}]$$

$$= \frac{a}{n}\left\{\frac{i(i-1)}{2} + \frac{(n-i)(n-i+1)}{2}\right\}$$

$$= \frac{a}{2n}\{2i^2 - 2(n+1)i + n(n+1)\}.$$

Then $E(Z) = E\{E(Z \mid X)\}$

$$= \sum_{i=1}^{n} E(Z \mid X=x_i)\, P(X=x_i)$$

$$= \sum_{i=1}^{n} \frac{a}{2n}\{2i^2 - 2(n+1)i + n(n+1)\}\frac{1}{n}$$

$$= \frac{a}{2n^2}\left\{2\sum_{i=1}^{n} i^2 - 2(n+1)\sum_{i=1}^{n} i + n^2(n+1)\right\}$$

$$= \frac{(n+1)(n-1)a}{3n}.$$

Ex. 20. *Let $U = X + aY$ and $V = X + \frac{\sigma_x}{\sigma_y} Y$, where a is a constant and σ_x, σ_y are the standard deviations of X, Y where X, Y are positively correlated. If $\rho(U, V) = 0$, then show that $a = -\frac{\sigma_x}{\sigma_y}$.*

$E(X + aY) = m_x + am_y$, where $m_x = E(X)$, $m_y = E(Y)$.

$$E\left(X + \frac{\sigma_x}{\sigma_y} Y\right) = m_x + \frac{\sigma_x}{\sigma_y} m_y.$$

The variances σ_u^2, σ_v^2 of U and V are given by

$$\sigma_u^2 = \text{var}\,(X + aY)$$

$$= E\{(X + aY) - (m_x + am_y)\}^2$$

$$= E\{(X - m_x) + a(Y - m_y)\}^2$$

$$\sigma_u^2 = E(X - m_x)^2 + a^2 E(Y - m_y)^2 + 2aE\{(X - m_x)(Y - m_y)\}$$

$$= \sigma_x^2 + a^2\sigma_y^2 + 2a\rho\,\sigma_x\sigma_y,$$

where ρ is the correlation coefficient between X and Y.

$$\sigma_v^2 = \text{var}\left(X + \frac{\sigma_x}{\sigma_y} Y\right)$$

$$= E\left\{(X - m_x) + \frac{\sigma_x}{\sigma_y}(Y - m_y)\right\}^2$$

$$= E(X - m_x)^2 + \frac{\sigma_x^2}{\sigma_y^2} E(Y - m_y)^2 + \frac{2\sigma_x}{\sigma_y} E\{(X - m_x)(Y - m_y)\}$$

$$= \sigma_x^2 + \frac{\sigma_x^2}{\sigma_y^2}\sigma_y^2 + 2\frac{\sigma_x}{\sigma_y}\rho\sigma_x\sigma_y$$

$$= 2(1 + \rho)\sigma_x^2.$$

$$\therefore \text{cov}\,(U, V) = E[\{(X - m_x) + a(Y - m_y)\}\{(X - m_x) + \frac{\sigma_x}{\sigma_y}(Y - m_y)\}]$$

$$= E\left\{(X - m_x)^2 + \left(a + \frac{\sigma_x}{\sigma_y}\right)(X - m_x)(Y - m_y) + \frac{a\sigma_x}{\sigma_y}(Y - m_y)^2\right\}$$

$$= E\{(X - m_x)^2\} + \left(a + \frac{\sigma_x}{\sigma_y}\right) E\{(X - m_x)(Y - m_y)\} + \frac{a\sigma_x}{\sigma_y} E\{(Y - m_y)^2\}$$

$$= \sigma_x^2 + \left(a + \frac{\sigma_x}{\sigma_y}\right)\rho\sigma_x\sigma_y + a\sigma_x\sigma_y$$

$$= (1 + \rho)\sigma_x^2 + a(1 + \rho)\sigma_x\sigma_y$$

$$\therefore \rho(U, V) = \frac{\text{cov}\,(U, V)}{\sigma_v\sigma_u} = 0 \text{ gives}$$

$$(1 + \rho)(\sigma_x^2 + a\sigma_x\sigma_y) = 0,$$

i.e., $a = -\frac{\sigma_x}{\sigma_y}$, $\because \sigma_x \neq 0, \sigma_y \neq 0$ and $\rho \neq -1$ since $\rho > 0$.

Ex. 21. *If $X_1, X_2, \ldots, X_n$ are mutually independent normal $(0, \sigma)$ variates, then show that*

$$\text{var}\left(\frac{X_1^2 + X_2^2 + \cdots\cdots + X_n^2}{n}\right) = \frac{2\sigma^4}{n}.$$

Here, $\text{var}\,X_i = \sigma^2$, $i = 1, 2, \ldots, n$.

$\therefore \sigma^2 = \text{var}\,X_i = E(X_i^2) - \{E(X_i)\}^2$.

But $E(X_i) = 0$ for $i = 1, 2, \ldots, n$.

So $E(X_i^2) = \sigma^2$ for $i = 1, 2, \ldots, n$.

Again, $X_1^2, X_2^2, \ldots, X_n^2$ are independent.

So, $\operatorname{var}\left(\dfrac{X_1^2+X_2^2+\cdots\cdots+X_n^2}{n}\right)$

$$=\frac{1}{n^2}(\operatorname{var} X_1^2+\operatorname{var} X_2^2+\cdots+\operatorname{var} X_n^2)$$

Now each X_i being normal $(0, \sigma)$,

$$\operatorname{var} X_i^2=E(X_i^4)-\{E(X_i^2)\}^2$$
$$=3\sigma^4-\sigma^4$$
$$=2\sigma^4, \text{ for } i=1, 2, \ldots\ldots, n.$$

$$\therefore \quad \operatorname{var}\left(\frac{X_1^2+X_2^2+\cdots+X_n^2}{n}\right)$$
$$=\frac{1}{n^2}\cdot n\cdot 2\sigma^4=\frac{2\sigma^4}{n}.$$

Ex. 22. *If X and Y are random variables, then show that no portion of the curve $y=\sigma_y^2x^2+2\rho\sigma_x\sigma_y x+\sigma_x^2$ can be below the x-axis, where σ_x, σ_y, ρ have the usual meanings.*

The equation $y=\sigma_y^2x^2+2\rho\sigma_x\sigma_y x+\sigma_x^2$ can be expressed as

$$y-\sigma_x^2=\sigma_y^2\left(x+\rho\frac{\sigma_x}{\sigma_y}\right)^2-\rho^2\sigma_x^2 \tag{8.10.6}$$

which is the equation of a parabola with vertex

$$\left\{-\rho\frac{\sigma_x}{\sigma_y}, (1-\rho^2)\sigma_x^2\right\},$$

a point lying on or above x-axis (since $-1\leqslant\rho\leqslant1$) and concavity upwards, axis parallel to the y-axis. Further we note that

$$y=\sigma_x^2(1-\rho^2)+\sigma_y^2\left(x+\rho\frac{\sigma_x}{\sigma_y}\right)^2\geqslant 0$$

for all real x.

Hence no portion of the the given curve can be below the x-axis.

Ex. 23. *The joint probability distribution of the discrete random variables X, Y is given by the following table :*

$Y \backslash X$	0	1
0	a	c
1	b	d

where a, b, c, d are non-negative real numbers. Show that X, Y are uncorrelated iff $ad=bc$. [C. H. (Math.) '83]

$$P(X=0)=a+b,\quad P(X=1)=c+d,$$
$$P(Y=0)=a+c,\quad P(Y=1)=b+d.$$
$$E(X)=0\,.\,(a+b)+1(c+d)=c+d=m_x,$$
$$E(Y)=0\,.\,(a+c)+1(b+d)=b+d=m_y.$$
$$E(X^2)=0\,.\,(a+b)+1^2(c+d)=c+d,$$
$$E(Y^2)=0\,.\,(a+c)+1^2(b+d)=b+d.$$
$$\therefore\quad \sigma_x^2=E(X^2)-\{E(X)\}^2=(c+d)-(c+d)^2=(c+d)(1-c-d)$$
and $\sigma_y^2=E(Y^2)-\{E(Y)\}^2=(b+d)-(b+d)^2=(b+d)(1-b-d)$.

Again, $E(XY)=0\times0\times a+0\times1\times c+1\times0\times b+1\times1\times d=d$.

$$\therefore\quad \operatorname{cov}(X, Y)=E(XY)-m_xm_y$$
$$=d-(c+d)(b+d).$$

X, Y are uncorrelated iff $\operatorname{cov}(X, Y)=0$,

i.e., iff $(c+d)(b+d)=d$,

i.e., iff $bc+d(b+c+d)=d$,

i.e., iff $bc+d(1-a)=d$, $\because\ a+b+c+d=1$,

i.e., iff $bc=ad$.

Ex. 24. *If X, Y are independent and Z, W are given by*

$$Z=X\cos\theta+Y\sin\theta,\ W=-X\sin\theta+Y\cos\theta,$$

then show that $\rho(Z, W)=\dfrac{(\sigma_y^2-\sigma_x^2)\sin 2\theta}{\sqrt{(\sigma_y^2-\sigma_x^2)\sin^2 2\theta+4\sigma_x^2\sigma_y^2}}.$

If m_z, m_w be the means of Z and W respectively,

$$m_z = E(X\cos\theta + Y\sin\theta) = m_x\cos\theta + m_y\sin\theta,$$

$$m_w = E(-X\sin\theta + Y\cos\theta) = -m_x\sin\theta + m_y\cos\theta,$$

where m_x and m_y are the means of X and Y respectively.

Also the variances σ_z^2 and σ_w^2 of Z and W are respectively given by

$$\begin{aligned}\sigma_z^2 &= E\{(Z-m_z)^2\}\\ &= E[\{(X-m_x)\cos\theta+(Y-m_y)\sin\theta\}^2]\\ &= \cos^2\theta\, E\{(X-m_x)^2\}+\sin^2\theta\, E\{(Y-m_y)^2\}\\ &\qquad +2\sin\theta\cos\theta\, E\{(X-m_x)(Y-m_y)\}\\ &= \sigma_x^2\cos^2\theta+\sigma_y^2\sin^2\theta,\end{aligned}$$

since X, Y being independent, $E\{(X-m_x)(Y-m_y)\}=\text{cov}\,(X, Y)=0$.

$$\begin{aligned}\sigma_w^2 &= E\{(W-m_w)^2\}\\ &= E[\{-(X-m_x)\sin\theta+(Y-m_y)\cos\theta\}^2]\\ &= \sin^2\theta\, E\{(X-m_x)^2\}+\cos^2\theta\, E\{(Y-m_y)^2\}\\ &\qquad -2\sin\theta\cos\theta\, E\{(X-m_x)(Y-m_y)\}\\ &= \sigma_x^2\sin^2\theta+\sigma_y^2\cos^2\theta.\end{aligned}$$

Again, cov (Z, W)

$$\begin{aligned}&= E[\{(X-m_x)\cos\theta+(Y-m_y)\sin\theta\}\{-(X-m_x)\sin\theta\\ &\qquad\qquad +(Y-m_y)\cos\theta\}]\\ &= E\{-\tfrac{1}{2}(X-m_x)^2\sin 2\theta+\tfrac{1}{2}(Y-m_y)^2\sin 2\theta\\ &\qquad +(X-m_x)(Y-m_y)\cos 2\theta\}\\ &= -\tfrac{1}{2}E\{(X-m_x)^2\}\sin 2\theta+\tfrac{1}{2}\{E(Y-m_y)^2\}\sin 2\theta\\ &\qquad +\cos 2\theta\, E\{(X-m_x)(Y-m_y)\}\\ &= -\tfrac{1}{2}(\sigma_x^2-\sigma_y^2)\sin 2\theta.\end{aligned}$$

$$\therefore\quad \rho(Z, W)=\frac{\text{cov}\,(Z, W)}{\sigma_z\sigma_w} =\frac{\tfrac{1}{2}(\sigma_y^2-\sigma_x^2)\sin 2\theta}{\sqrt{(\sigma_x^2\cos^2\theta+\sigma_y^2\sin^2\theta)(\sigma_x^2\sin^2\theta+\sigma_y^2\cos^2\theta)}}.$$

Now, $(\sigma_x^2\sin^2\theta+\sigma_y^2\cos^2\theta)(\sigma_x^2\cos^2\theta+\sigma_y^2\sin^2\theta)$

$$\begin{aligned}&=\sigma_x^4\sin^2\theta\cos^2\theta+\sigma_y^4\sin^2\theta\cos^2\theta+\sigma_x^2\sigma_y^2(\cos^4\theta+\sin^4\theta)\\ &=(\sigma_x^2-\sigma_y^2)^2\sin^2\theta\cos^2\theta+\sigma_x^2\sigma_y^2(\cos^4\theta+\sin^4\theta+2\sin^2\theta\cos^2\theta)\\ &=\tfrac{1}{4}(\sigma_x^2-\sigma_y^2)^2\sin^2 2\theta+\sigma_x^2\sigma_y^2\\ &=\tfrac{1}{4}\{(\sigma_x^2-\sigma_y^2)^2\sin^2 2\theta+4\sigma_x^2\sigma_y^2\}.\end{aligned}$$

$$\therefore\quad \rho(Z, W)=\frac{(\sigma_y^2-\sigma_x^2)\sin 2\theta}{\sqrt{(\sigma_x^2-\sigma_y^2)^2\sin^2 2\theta+4\sigma_x^2\sigma_y^2}}.$$

Ex. 25. *Find the expectation of the sum of points on n unbiased dice.*

Let the random variable X_i denote the number of points on the ith die ($i=1, 2, \ldots, n$). Then the random variable S denoting the sum of points on n dice is given by

$$S=X_1+X_2+\cdots\cdots+X_n.$$

$$\begin{aligned}\therefore\quad E(S)&=E(X_1+X_2+\cdots\cdots+X_n)\\ &=E(X_1)+E(X_2)+\cdots+E(X_n).\end{aligned}$$

Now each of the random variable X_i assumes values 1, 2, 3, 4, 5, 6 each having the probability $\frac{1}{6}$.

$$\begin{aligned}\therefore\quad E(X_i)&=1\times\tfrac{1}{6}+2\times\tfrac{1}{6}+3\times\tfrac{1}{6}+4\times\tfrac{1}{6}+5\times\tfrac{1}{6}+6\times\tfrac{1}{6}\\ &=\tfrac{7}{2},\ \text{for } i=1, 2, \ldots\ldots, n.\end{aligned}$$

$$\therefore\quad E(S)=\frac{7n}{2}.$$

Ex. 26. *If X, Y be independent random variables with means m_x, m_y and variances σ_x^2, σ_y^2 respectively, then prove that*

$$\text{var}\,(XY)=\sigma_x^2\sigma_y^2+m_x^2\sigma_y^2+m_y^2\sigma_x^2.$$

$E(XY)=E(X)\,E(Y)=m_x m_y$, X and Y being independent.

Now var (XY)

$$\begin{aligned}&=E\,(XY-m_x m_y)^2\\ &=E\,\{(X-m_x)(Y-m_y)-2m_x m_y+m_x Y+m_y X\}^2\\ &=E\,\{(X-m_x)(Y-m_y)+m_x\,(Y-m_y)+m_y\,(X-m_x)\}^2\\ &=E\,\{(X-m_x)^2\,(Y-m_y)^2\}+m_x^2\,E\,\{(Y-m_y)^2\}\\ &\qquad +m_y^2\,E\,\{(X-m_x)^2\}+2m_x\,E\,\{(X-m_x)(Y-m_y)^2\}\\ &\qquad +2m_y\,E\,\{(X-m_x)^2\,(Y-m_y)\}\\ &\qquad +2m_x m_y\,E\,\{(X-m_x)(Y-m_y)\}\\ &=E\,\{(X-m_x)^2\}\,E\,\{(Y-m_y)^2\}+m_x^2\,E\,\{(Y-m_y)^2\}\\ &\qquad +m_y^2\,E\,\{(X-m_x)^2\}+2m_x\,E(X-m_x)\,E\,\{(Y-m_y)^2\}\\ &\qquad +2m_y\,E\,\{(X-m_x)^2\}\,E(Y-m_y)\\ &\qquad +2m_x m_y\,E\,(X-m_x)\,E\,(Y-m_y)\\ &=\sigma_x^2\sigma_y^2+m_x^2\sigma_y^2+m_y^2\sigma_x^2,\end{aligned}$$

since, X and Y being independent,

$$E\,\{(X-m_x)^2\,(Y-m_y)^2\}=E\,\{(X-m_x)^2\}\,E\,\{(Y-m_y)^2\},$$

$$E\,\{(X-m_x)(Y-m_y)\}=E\,(X-m_x)\,E\,(Y-m_y)$$

and also $E\,(X-m_x)=E\,(Y-m_y)=0.$

Ex. 27. *The joint distribution of the discrete random variables X, Y is given by*

Y \ X	0	1
0	p_{00}	p_{10}
1	p_{01}	p_{11}

Find the characteristic functions $\phi_X(t_1)$, $\phi_Y(t_2)$, $\phi_{X,Y}(t_1, t_2)$ and show that $\phi_{X,Y}(t_1, t_2)=\phi_X(t_1)\,\phi_Y(t_2)$ if $\begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11}\end{bmatrix}$ is a singular matrix.

The characteristic function $\phi_{X,Y}(t_1, t_2)$ of the joint distribution of X, Y is given by

$$\phi_{X,Y}(t_1, t_2)=E\left\{e^{i(t_1X+t_2Y)}\right\}$$

$$=p_{00}\,e^{i0}+p_{10}\,e^{it_1}+p_{01}e^{it_2}+p_{11}e^{i(t_1+t_2)}$$

$$=p_{00}+p_{10}\,e^{it_1}+p_{01}e^{it_2}+p_{11}e^{i(t_1+t_2)}.$$

Again, $\phi_X(t_1)=(p_{00}+p_{01})e^{i0}+(p_{10}+p_{11})e^{it_1}$

$$=p_{00}+p_{01}+(p_{10}+p_{11})e^{it_1}=\phi_{X,Y}(t_1, 0)$$

and $\phi_Y(t_2)=(p_{00}+p_{10})\,e^{i0}+(p_{01}+p_{11})e^{it_2}$

$$=p_{00}+p_{10}+(p_{01}+p_{11})\,e^{it_2}=\phi_{X,Y}(0, t_2).$$

Now let $\begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11}\end{bmatrix}$ be singular. Then

$$\begin{vmatrix} p_{00} & p_{10} \\ p_{01} & p_{11}\end{vmatrix}=0, \quad i.e., \quad p_{00}\,p_{11}=p_{01}\,p_{10}.$$

Case I. $p_{11}\neq 0$.

Here $\phi_{X,Y}(t_1, t_2)$

$$=p_{11}^{-1}\{p_{00}p_{11}+p_{11}p_{10}e^{it_1}+p_{11}p_{01}e^{it_2}+p_{11}^2e^{i(t_1+t_2)}\}$$

$$=p_{11}^{-1}(p_{01}+p_{11}e^{it_1})(p_{10}+p_{11}e^{it_2}), \quad (\because\ p_{00}p_{11}=p_{01}p_{10}).$$

Also $\phi_X(t_1)=\phi_{X,Y}(t_1, 0)=p_{11}^{-1}(p_{10}+p_{11})(p_{01}+p_{11}e^{it_1})$

and $\phi_Y(t_2)=\phi_{X,Y}(0, t_2)=p_{11}^{-1}(p_{01}+p_{11})(p_{10}+p_{11}e^{it_2})$.

Hence, $\phi_{X,Y}(t_1, t_2)=\phi_X(t_1)\,\phi_Y(t_2)$ if

$$p_{11}^{-1}(p_{01}+p_{11}e^{it_1})(p_{10}+p_{11}{}^{it_2})$$

$$=p_{11}^{-1}(p_{10}+p_{11})(p_{01}+p_{11}e^{it_1})p_{11}^{-1}(p_{01}+p_{11})(p_{10}+p_{11}e^{it_2}),$$

i.e., if $p_{11}^{-1}(p_{10}+p_{11})(p_{01}+p_{11})=1$,

i.e., if $p_{10}p_{01}+p_{10}p_{11}+p_{11}p_{01}+p_{11}^2=p_{11}$,

i.e., if $p_{00}p_{11}+p_{10}p_{11}+p_{11}p_{01}+p_{11}^2=p_{11}$, $\quad \because\ p_{00}p_{11}=p_{01}p_{10}$,

i.e., if $p_{00}+p_{10}+p_{01}+p_{11}=1$, which is true.

Case II. $p_{11}=0$.

Since $p_{11}p_{00}=p_{01}p_{10}$, we get $p_{01}=0$ or $p_{10}=0$.

If $p_{10}=0$, then $p_{10}=p_{11}=0$ and $p_{01}+p_{00}=1$.

Then $\phi_{X,Y}(t_1, t_2)=p_{00}+p_{01}e^{it_2}$,

and $\phi_X(t_1)=p_{00}+p_{01}=1$,

$$\phi_Y(t_2)=p_{00}+p_{01}e^{it_2}.$$

So, $\phi_{X,Y}(t_1, t_2)=\phi_X(t_1)\,\phi_Y(t_2)$ if $p_{10}=0$.

Similarly, we can show that

$$\phi_{X,Y}(t_1, t_2)=\phi_X(t_1)\,\phi_Y(t_2) \quad \text{if} \quad p_{01}=0.$$

Hence, it is proved that in any case

$$\phi_{X,Y}(t_1, t_2)=\phi_X(t_1)\,\phi_Y(t_2)$$

if $\begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11}\end{bmatrix}$ is a singular matrix.

Ex. 28. *Let $X_1, X_2, \ldots\ldots, X_n$ be independent random variables. Show that if k be any positive integer less than n and*

$$Y_k=a_1X_1+a_2X_2+\ldots\ldots+a_kX_k$$

where a_i's are constants, then Y_k, X_{k+1} are independent.

[*C. H. (Math.)* '76]

The joint characteristic function of Y_k and X_{k+1} is given by

$$\phi_{Y_k,X_{k+1}}(t, u)=E\{e^{i(tY_k+uX_{k+1})}\}$$

$$=E\{e^{ita_1X_1}\cdot e^{ita_2X_2}\ldots e^{ita_kX_k}\cdot e^{iuX_{k+1}}\}.$$

Since $X_1, X_2, \ldots\ldots, X_n$ are mutually independent and k is a positive integer $<n$, $X_1, X_2, \ldots\ldots, X_k, X_{k+1}$ are also mutually independent.

Hence, $\phi_{Y_k, X_{k+1}}(t, u)$

$$=E(e^{ita_1X_1}) \cdot E(e^{ita_2X_2}) \ldots E(e^{ita_kX_k}) \cdot E(e^{iuX_{k+1}})$$

$$=E(e^{ita_1X_1} \cdot e^{ita_2X_2} \cdot \ldots e^{ita_kX_k})\, E(e^{iuX_{k+1}}),$$

since $X_1, X_2, \ldots, X_k$ are mutually independent

$$=E\{e^{it(a_1X_1+a_2X_2+\ldots+a_kX_k)}\} \cdot E(e^{iuX_{k+1}})$$

$$=E(e^{itY_k})\, E(e^{iuX_{k+1}})$$

$$=\phi_{Y_k}(t)\ \phi_{X_{k+1}}(u),$$

where $\phi_{Y_k}(t), \phi_{X_{k+1}}(u)$ are the characteristic functions of Y_k and X_{k+1} respectively. Thus we get

$$\phi_{Y_k, X_{k+1}}(t, u)=\phi_{Y_k}(t)\ \phi_{X_{k+1}}(u).$$

So, by converse of Theorem 8.5.2, Y_k and X_{k+1} are independent.

Ex. 29. *The joint probability density function of the random variables X and Y is $\frac{1}{2}x^3 e^{-x(y+1)}$ $(0<x<\infty, 0<y<\infty)$. Determine the correlation ratio of Y on X.* [C. H. (Math.) '91]

The correlation ratio of Y on X is the correlation coefficient between the random variables $m_Y(X)$ and Y, where $m_Y(X)$ is the random variable corresponding to $m_Y(x)$ and $m_Y(x)=E(Y \mid X=x)$ $=\int_{-\infty}^{\infty} y\,\frac{f(x,y)}{f_X(x)}\,dy$, $f(x,y)$ is the joint probability density function of X and Y and $f_X(x)$ is the marginal probability density function of X.

$$\text{Here } f_X(x)=\int_0^{\infty} \tfrac{1}{2}x^3 e^{-x(y+1)}\,dy=\tfrac{1}{2}x^3\int_0^{\infty} e^{-x(y+1)}dy$$

$$=\tfrac{1}{2}x^3 e^{-x}\int_0^{\infty} e^{-xy}dy=\tfrac{1}{2}x^2 e^{-x},\ 0<x<\infty.$$

$$\text{So, } m_Y(x)=\int_0^{\infty}\frac{y \cdot \frac{1}{2}x^3 e^{-x(y+1)}}{\frac{1}{2}x^2 e^{-x}}\,dy$$

$$=\int_0^{\infty} xy\, e^{-xy}\,dy$$

$$=\frac{1}{x}\int_0^{\infty} ze^{-z}\,dz, \quad \text{where } z=xy$$

$$=\frac{1}{x}\,\Gamma(2)=\frac{1}{x}.$$

$$\therefore\ m_Y(X)=\frac{1}{X}.$$

Then the required correlation ratio

$$=\rho\,\{m_Y(X), Y\}$$

$$=\frac{\operatorname{cov}\left(\frac{1}{X}, Y\right)}{\sigma_1\sigma_2}$$

where σ_1, σ_2 are the standard deviations of $\frac{1}{X}$ and Y respectively.

$$\text{Now, } E\left(\frac{1}{X}\right)=\int_0^{\infty}\int_0^{\infty}\frac{1}{x}\cdot\tfrac{1}{2}\cdot x^3 e^{-x(y+1)}\,dx\,dy$$

$$=\frac{1}{2}\int_0^{\infty}\left\{x^2\int_0^{\infty} e^{-x(y+1)}\,dy\right\}dx$$

$$=\frac{1}{2}\int_0^{\infty} xe^{-x}\,dx=\frac{\Gamma(2)}{2}=\tfrac{1}{2},$$

$$E(Y)=m_y=\int_0^{\infty}\int_0^{\infty} y\cdot\tfrac{1}{2}x^3 e^{-x(y+1)}\,dx\,dy$$

$$=\frac{1}{2}\int_0^{\infty}\left(x^3e^{-x}\int_0^{\infty} ye^{-xy}dy\right)dx$$

$$=\frac{1}{2}\int_0^{\infty} xe^{-x}dx=\tfrac{1}{2}.$$

$$\therefore \quad \sigma_1^2 = E\left(\frac{1}{X^2}\right) - \frac{1}{4} = \int_0^\infty \frac{1}{x^2} f_X(x)\, dx - \frac{1}{4}$$

$$= \int_0^\infty \frac{1}{x^2} \cdot \frac{1}{2} x^2 e^{-x} dx - \frac{1}{4}$$

$$= \frac{1}{2} - \frac{1}{4} = \frac{1}{4},$$

$$\sigma_2^2 = E(Y^2) - \frac{1}{4}.$$

Now,
$$E(Y^2) = \int_0^\infty \int_0^\infty y^2 \cdot \frac{1}{2} x^3 e^{-x(y+1)}\, dx\, dy$$

$$= \int_0^\infty \frac{1}{2} x^3 e^{-x} \left(\int_0^\infty y^2 e^{-xy}\, dy \right) dx$$

$$= \int_0^\infty \frac{1}{2} x^3 e^{-x} \left(\frac{1}{x^3} \int_0^\infty z^2 e^{-z}\, dz \right) dx, \text{ where } xy = z$$

$$= \frac{1}{2} \Gamma(3) \int_0^\infty e^{-x}\, dx = 1.$$

$$\therefore \quad \sigma_2^2 = 1 - \frac{1}{4} = \frac{3}{4}.$$

Now,
$$\operatorname{cov}\left(\frac{1}{X}, Y\right) = E\left(\frac{Y}{X}\right) - E\left(\frac{1}{X}\right) E(Y)$$

$$= E\left(\frac{Y}{X}\right) - \frac{1}{4}.$$

Again,
$$E\left(\frac{Y}{X}\right) = \int_0^\infty \int_0^\infty \frac{y}{x} \frac{1}{2} x^3 e^{-x(y+1)}\, dx\, dy$$

$$= \int_0^\infty \left(\frac{1}{2} x^2 e^{-x} \int_0^\infty y e^{-xy}\, dy \right) dx$$

$$= \frac{1}{2} \int_0^\infty e^{-x}\, dx = \frac{1}{2}.$$

$$\therefore \quad \rho\{ m_Y(X), Y \} = \frac{\frac{1}{2} - \frac{1}{4}}{\frac{1}{2} \cdot \frac{\sqrt{3}}{2}} = \frac{1}{\sqrt{3}}.$$

Ex. 30. *If $g: R \to R$ be a continuous function, then prove that $\rho\{ g(X), Y - m_Y(X) \} = 0$, where the joint distribution of X and Y is continuous.*

We have

$\operatorname{cov}\{ g(X), Y - m_Y(X) \} = E[\, g(X)\{Y - m_Y(X)\}\,] - m_1 m_2$,

where $m_1 = E\{g(X)\}$, $m_2 = E\{Y - m_Y(X)\}$.

Now,
$$m_2 = E\{Y - m_Y(X)\}$$
$$= E(Y) - E\{m_Y(X)\}$$
$$= E(Y) - E\{E(Y \mid X)\}$$
$$= E(Y) - E(Y), \text{ by } (8.6.18)$$
$$= 0.$$

Also $E[g(X)\{Y - m_Y(X)\}]$

$$= E\{Y g(X)\} - E\{g(X)\, m_Y(X)\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, g(x)\, f(x, y)\, dx\, dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\{E(Y \mid X = x\} f_X(x)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, g(x)\, f(x, y)\, dx\, dy - \int_{-\infty}^{\infty} g(x) \left\{ \int_{-\infty}^{\infty} y \frac{f(x, y)}{f_X(x)} dy \right\} f_X(x) dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, g(x)\, f(x, y)\, dx\, dy - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\, g(x)\, f(x, y)\, dx\, dy$$

$$= 0,$$

where, $f(x, y)$, $f_X(x)$ have the usual meanings.

Hence, $\operatorname{cov}\{ g(X), Y - m_Y(X) \} = 0$ and so

$$\rho\{ g(X), Y - m_Y(X) \} = 0.$$

Ex. 31. *Using the result of Ex. 30, prove that*

$$E[\{m_Y(X) - c_0^* - c_1^* X\}^2] = \sigma_y^2 (\eta^2 - \rho^2)$$

where $y = c_0^* + c_1^* x$ is the least square regression line of Y on X, ρ is the correlation coefficint of X and Y and η is the correlation ratio of Y on X.

[Joint distribution of X and Y is continuous.]

Since $y = c_0^* + c_1^* x$ is the least square regression line of Y on X, we have, by (8.8.9),

$$c_0^* = m_y - \rho \frac{\sigma_y}{\sigma_x} m_x, \quad c_1^* = \rho \frac{\sigma_y}{\sigma_x}.$$

We have

$$E\{(Y-c_0^*-c_1^*X)^2\}$$
$$=E[\{Y-m_Y(X)+m_Y(X)-c_0^*-c_1^*X\}^2]$$
$$=E[\{Y-m_Y(X)\}^2]+E[\{m_Y(X)-c_0^*-c_1^*X\}^2]$$
$$+2E[\{Y-m_Y(X)\}\{m_Y(X)-c_0^*-c_1^*X\}].$$

Now, $\text{cov}\{Y-m_Y(X),\ m_Y(X)-c_0^*-c_1^*X\}$

$$=E[\{Y-m_Y(X)\}\{m_Y(X)-c_0^*-c_1^*X\}]$$
$$-E[\{Y-m_Y(X)\}]\ E[\{m_Y(X)-c_0^*-c_1^*X\}].$$

Since $E\{Y-m_Y(X)\}=m_y-m_y=0$,

and $E\{m_Y(X)-c_0^*-c_1^*X\}$

$$=m_y-c_0^*-c_1^*m_x$$
$$=m_y-m_y+\rho\frac{\sigma_y}{\sigma_x}m_x-\rho\frac{\sigma_y}{\sigma_x}m_x$$
$$=0,$$

$\therefore\ E\{(Y-c_0^*-c_1^*X)^2\}$

$$=E[\{Y-m_Y(X)\}^2]+E[\{m_Y(X)-c_0^*-c_1^*X\}^2]$$
$$+2\,\text{cov}\{Y-m_Y(X),\ m_Y(X)-c_0^*-c_1^*X\}.$$

Now by Ex. 30, taking $g(X)=m_Y(X)-c_0^*-c_1^*X$, we find that $\text{cov}\{m_Y(X)-c_0^*-c_1^*X,\ Y-m_Y(X)\}=0$.

Hence, $E\{(Y-c_0^*-c_1^*X)^2\}$

$$=E[\{Y-m_Y(X)\}^2]+E[\{m_Y(X)-c_0^*-c_1^*X\}^2]. \tag{8.10.7}$$

Now, by (8.8.11), we have

$$E\{(Y-c_0^*-c_1^*X)^2\}=\sigma_y^2(1-\rho^2).$$

Again, we know that $y=m_Y(x)$ is the least square regression curve of Y on X from the family of all curves of $y=g(x)$ ($g(x)$ is continuous). Then taking $u=m_Y(x)$, $y=u$ can be taken as the least square regression line of Y on U from the family of straight lines $y=a+bu$ when u is the real variable corresponding to the random variable $U=m_Y(X)$. Then $a^*=0$, $b^*=1$. So comparing with (8.8.11)

$$E\{(Y-a^*-b^*U)^2\}=\sigma_y^2[1-\{\rho(U,Y)\}^2]$$

or, $E[\{Y-m_Y(X)\}^2]=\sigma_y^2[1-\{\rho(m_Y,X),\ Y)\}^2]$

or, $E[\{Y-m_Y(X)\}^2]=\sigma_y^2(1-\eta^2)$.

Hence, from (8.10.7), we get

$$\sigma_y^2(1-\rho^2)=\sigma_y^2(1-\eta^2)+E[\{m_Y(X)-c_0^*-c_1^*X\}^2],$$

$$\therefore\ E[\{m_Y(X)-c_0^*-c_1^*X\}^2]=\sigma_y^2(\eta^2-\rho^2).$$

Ex. 32. *Let X_1 be the random variable denoting the number of trials up to and including the first success and X_k be the random variable denoting the number of trials following the $(k-1)$th success up to and including the kth success (for $k=2, 3, 4,\ldots$), in an infinite sequence of Bernoulli's trials where p is the probability of success in each trial. Prove that*

$$\text{var}\,(X_1+X_2+\cdots+X_n)=\frac{n(1-p)}{p^2}.$$

Further if X be the random variable denoting the run of successes or failures starting with the first trial, then show that

$$\text{var}\ X=\frac{q}{p}+\frac{p}{q}+\left(\frac{p}{q}-\frac{q}{p}\right)^2$$

where $q=1-p$.

Here $X_1, X_2,\ldots\ldots, X_n$ all have the spectrum $\{1, 2, 3,\ldots\ldots\}$.

Then $E(X_k)=\sum_{r=1}^{\infty} rP(X_k=r)$.

Now, $X_k=r$ denotes the event 'success follows $r-1$ consecutive failures immediately after the $(k-1)$th success' for $k=2, 3\ldots$ and $X_1=r$ denotes the event 'success follows $r-1$ consecutive failures starting from the first trial'.

Then $P(X_k=r)=(1-p)^{r-1}p$.

So $E(X_k)=\sum_{r=1}^{\infty} r(1-p)^{r-1}p$

$$=p\{1+2(1-p)+3(1-p)^2+\cdots\}$$
$$=p\{1-(1-p)\}^{-2}$$
$$=\frac{1}{p}.$$

Thus we have

$$E(X_1)=E(X_2)=\cdots=E(X_n)=\frac{1}{p}.$$

$$\therefore\ E(X_1+X_2+\cdots+X_n)=\frac{n}{p}.$$

Now, let $Y=X_1+X_2+\cdots+X_n$. Then Y is the random variable denoting the total number of trials up to and including the nth success. So the spectrum of Y is the set $\{n, n+1, n+2, \ldots\}$.

Now, $Y=n$ denotes the event 'n successes in the first n trials'.

$\therefore \quad P(Y=n)=p^n.$

$Y=n+1$ denotes the event '$n-1$ successes in the first n trials and success in the $(n+1)$th trial'.

$$\therefore \quad P(Y=n+1)={}^n c_{n-1}\, p^{n-1}(1-p)\cdot p$$
$$={}^n c_1\, p^n(1-p).$$

In general, we have

$$P(Y=n+r)={}^{n+r-1}c_r\, p^n(1-p)^r \; ; \; r=0, 1, 2, \ldots .$$

$$\text{So, } E(Y^2)=\sum_{r=0}^{\infty}(n+r)^2\, {}^{n+r-1}c_r\, p^n(1-p)^r$$

$$=p^n\{n^2+(n+1)^2\; {}^n c_1(1-p)+(n+2)^2\; {}^{n+1}c_2(1-p)^2+\cdots\}$$

$$=p^n\left\{n^2+(n+1)^2\, n(1-p)+\frac{(n+2)^2(n+1)n}{2}(1-p)^2+\cdots\right\}$$

$$=np^n\left\{n+(n+1)\cdot(n+1)(1-p)+(n+2)\cdot\frac{(n+2)(n+1)}{2\,!}(1-p)^2+\cdots\right\}.$$

Now, we know that

$$(1-x)^{-(n+1)}=1+(n+1)x+\frac{(n+1)(n+2)}{2\,!}x^2+\cdots, \quad \text{if } |x|<1.$$

$$\therefore \quad x^n(1-x)^{-(n+1)}=x^n+(n+1)\,x^{n+1}+\frac{(n+1)(n+2)}{2\,!}x^{n+2}+\cdots$$

Here the process of term by term differentiation is valid.

$$\therefore \quad \frac{d}{dx}\left\{\frac{x^n}{(1-x)^{n+1}}\right\}=nx^{n-1}+(n+1)(n+1)x^n+\frac{(n+1)(n+2)(n+2)}{2\,!}x^{n+1}+\cdots$$

$$=x^{n-1}\left\{n+(n+1)(n+1)x+(n+2)\frac{(n+2)(n+1)}{2\,!}x^2+\cdots\right\}$$

or,
$$\frac{nx^{n-1}(1-x)^{n+1}+(n+1)(1-x)^n\cdot x^n}{(1-x)^{2n+2}}$$
$$=x^{n-1}\left\{n+(n+1)(n+1)\,x+(n+2)\cdot\frac{(n+2)(n+1)}{2\,!}x^2+\cdots\right\} \tag{8.10.8}$$

We have $0<1-p<1$. So we can take $x=1-p$ in (8.10.8). Then we get

$$n+(n+1)(n+1)(1-p)+(n+2)\frac{(n+2)(n+1)}{2\,!}(1-p)^2+\cdots$$
$$=\frac{np^{n+1}+(n+1)\,p^n(1-p)}{p^{2n+2}}.$$

$$\therefore \quad E(Y^2)=np^n\cdot\frac{np^{n+1}+(n+1)\,p^n(1-p)}{p^{2n+2}}$$
$$=\frac{n^2p+n(n+1)(1-p)}{p^2}$$
$$=\frac{n(n+1)-np}{p^2}.$$

$$\therefore \quad \text{var } Y=E(Y^2)-\{E(Y)\}^2$$
$$=\frac{n(n+1)-np}{p^2}-\frac{n^2}{p^2}$$
$$\left[\because \; E(Y)=E(X_1+\cdots+X_n)=\frac{n}{p}\right]$$
$$=\frac{n(1-p)}{p^2}.$$

So, it is proved that

$$\text{var } (X_1+X_2+\cdots+X_n)=\frac{n(1-p)}{p^2}.$$

Second Part: Here $X=r$ denotes the event A_r+B_r where A_r is the event 'success occurs in each of the first r trials and failure in the $(r+1)$th trial', B_r is the event 'failure occurs in each of the first r trials and success occurs in the $(r+1)$th trial'. We note that A_r, B_r are mutually exclusive.

Then $P(X=r)=P(A_r)+P(B_r)$.

Now, $P(A_r)=p^r\cdot q$ and $P(B_r)=q^r\cdot p$.

So, $P(X=r)=p^rq+q^rp, \quad r=1, 2, \ldots .$

$$\text{Then } E(X)=\sum_{r=1}^{\infty} r\,(p^rq+q^rp)$$
$$=q\cdot\sum_{r=1}^{\infty} rp^r+p\sum_{r=1}^{\infty} rq^r$$
$$=q\,(p+2p^2+3p^3+\cdots)+p\,(q+2q^2+3q^3+\cdots$$
$$=qp\,(1-p)^{-2}+pq(1-q)^{-2}$$
$$=\frac{p}{q}+\frac{q}{p}.$$

$$E(X^2)=q\sum_{r=1}^{\infty} r^2p^r+p\sum_{r=1}^{\infty} r^2q^r.$$

Now, we have $(1-x)^{-2}=1+2x+3x^2+4x^3+\cdots$, if $|x|<1$.

$$\therefore \quad \frac{x}{(1-x)^2}=x+2x^2+3x^3+4x^4+\cdots.$$

Then $\dfrac{d}{dx}\left\{\dfrac{x}{(1-x)^2}\right\}=1+2^2x+3^2x^2+4^2x^3+\cdots$

$$\therefore \quad x\frac{d}{dx}\left\{\frac{x}{(1-x)^2}\right\}=x+2^2x^2+3^2x^3+4^2x^4+\cdots$$

or, $$x\,\frac{(1-x)^2+2(1-x)x}{(1-x)^4}=x+2^2x^2+3^2x^3+4^2x^4+\cdots. \quad (8.10.9)$$

Taking $x=p$ in (8.10.9) $(0<p<1)$ we get

$$\frac{p(1+p)}{(1-p)^3}=p+2^2p^2+3^2p^3+4^2p^4+\cdots$$

$$\therefore \quad \sum_{r=1}^{\infty} p^r r^2=\frac{p(1+p)}{q^3}.$$

i.e., $$q\sum_{r=1}^{\infty} p^r r^2=\frac{p(1+p)}{q^2}.$$

Similarly, $$p\sum_{r=1}^{\infty} q^r r^2=\frac{q(1+q)}{p^2}.$$

Hence, $$E(X^2)=\frac{p(1+p)}{q^2}+\frac{q(1+q)}{p^2}.$$

So $\operatorname{var} X=E(X^2)-\{E(X)\}^2$

$$=\frac{p}{q^2}+\frac{p^2}{q^2}+\frac{q}{p^2}+\frac{q^2}{p^2}-\left(\frac{p}{q}+\frac{q}{p}\right)^2$$

$$=\frac{p}{q^2}+\frac{q}{p^2}-2$$

$$=\frac{p^3+q^3}{p^2q^2}-2$$

$$=\frac{(p^3+q^3)(p+q)}{p^2q^2}-2, \quad \because\ p+q=1.$$

or, $$\operatorname{var} X=\frac{p^3+q^3}{pq^2}+\frac{p^3+q^3}{p^2q}-2$$

$$=\frac{p^2}{q^2}+\frac{q}{p}+\frac{p}{q}+\frac{q^2}{p^2}-2$$

$$=\left(\frac{p^2}{q^2}+\frac{q^2}{p^2}-2\right)+\left(\frac{q}{p}+\frac{p}{q}\right)$$

$$=\left(\frac{p}{q}-\frac{q}{p}\right)^2+\frac{q}{p}+\frac{p}{q}.$$

Examples VIII

1. The joint probability density function of the random variables X and Y is given by

$$f(x, y)=\frac{2}{a^2} \text{ if } 0\leqslant y<x,\ 0\leqslant x\leqslant a.$$

Find the conditional mean of Y on X. If $y=2x$ and $x=\frac{y}{8}$ are the two regression lines for a given bivariate distribution, find $\rho(X, Y)$. Also obtain the regression line of U and V, where $U=X+Y$, $V=X-Y$. [C. H. (Math.) '68]

2. Find $\rho(X, Y)$ (i) between X and $Y=3-4X$, (ii) between X and $Y=X^2$, where X is a normal variate with zero mean. [C. H. (Math.) '71]

3. The joint probability density function of X and Y is

$$f(x, y)=8\,xy \text{ if } 0\leqslant x\leqslant y,\ 0\leqslant y\leqslant 1$$
$$=0 \quad \text{elsewhere.}$$

Examine whether X and Y are independent. Also compute var X and var Y. [C. H. (Math.) '88]

[*Hint* : $f_X(x)=\displaystyle\int_x^1 8xy\,dy=4x\,(1-x^2),\ 0\leqslant x\leqslant 1,$

$f_Y(y)=\displaystyle\int_0^y 8xy\,dx=4y^3,\ 0\leqslant y\leqslant 1,$

$f(x, y)\neq f_X(x)\,f_Y(y)$. So X, Y are not independent.

$$E(X^2)=\int_0^1 x^2 f_X(x)\,dx=\int_0^1 x^2\,.\,4x(1-x^2)\,dx=\tfrac{1}{3}.$$

$$E(X)=\int_0^1 4x^2(1-x^2)\,dx=\tfrac{8}{15}.$$

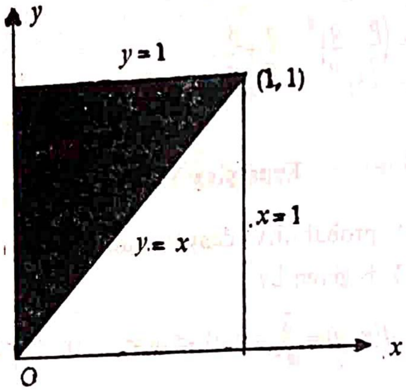


Fig. 8.10.3

$$\therefore \quad \text{var } X=E(X^2)-\{E(X)\}^2=\tfrac{1}{3}-(\tfrac{8}{15})^2=\tfrac{11}{225},$$

$$\text{var } Y=E(Y^2)-\{E(Y)\}^2$$

$$=\int_0^1 4y^5\,dy-\left(\int_0^1 4y^4\,dy\right)^2$$

$$=\tfrac{2}{3}-(\tfrac{4}{5})^2$$

$$=\tfrac{2}{75}.\ ]$$

4. If the joint probability density function of X and Y be $f(x, y)=a^2 e^{-ay}, 0 \leqslant x \leqslant y, 0 \leqslant y < \infty$, then find the value of $\rho(X, Y)$ $(a > 0)$.

[*Hint* : Here $\int_0^\infty\left(\int_x^\infty a^2e^{-ay}\,dy\right)dx=1$ for any $a > 0$,

$$f_X(x)=\int_x^\infty a^2 e^{-ay}\,dy=ae^{-ax},\ 0 \leqslant x < \infty,$$

$$f_Y(y)=\int_0^y a^2 e^{-ay}\,dx=a^2 ye^{-ay}.\ 0 \leqslant y < \infty.$$

$$E(X)=\int_0^\infty ax\,e^{-ax}\,dx=\frac{1}{a},$$

$$E(X^2)=\int_0^\infty ax^2\,e^{-ax}\,dx=\frac{2}{a^2},$$

$$E(Y)=\int_0^\infty a^2y^2\,e^{-ay}\,dy=\frac{2}{a},$$

$$E(Y^2)=\int_0^\infty a^2y^3\,e^{-ay}\,dy=\frac{6}{a^2}.$$

$$\text{var } X=\frac{2}{a^2}-\frac{1}{a^2}=\frac{1}{a^2}$$

$$\text{var } Y=\frac{6}{a^2}-\frac{4}{a^2}=\frac{2}{a^2},$$

$$\text{cov}(X, Y)=E(XY)-E(X)\,E(Y)$$

$$=E(XY)-\frac{2}{a^2}.$$

$$E(XY)=\int_0^\infty\left(\int_x^\infty xy\,a^2e^{-ay}\,dy\right)dx$$

$$=\int_0^\infty\left(x\int_{ax}^\infty ze^{-z}\,dz\right)dx \quad \text{where } ay=z$$

$$=\int_0^\infty x\left\{\underset{B\to\infty}{Lt}\left(-Be^{-B}-e^{-B}+ax\,e^{-ax}+e^{-ax}\right)\right\}dx$$

$$=\int_0^\infty (ax^2\,e^{-ax}+xe^{-ax})\,dx$$

$$=\frac{2}{a^2}+\frac{1}{a^2}=\frac{3}{a^2}.$$

$$\therefore \quad \rho(X,Y)=\frac{\text{cov}(X, Y)}{\sigma_x\sigma_y}=\frac{1}{\sqrt{2}}.\ ]$$

5. Show that $\rho(X, Y)=-\frac{1}{3}$ if the joint probability density function of X and Y is given by

$$f(x, y)=\tfrac{1}{2}\,x^3\,e^{-x(y+1)}\,;\ x > 0, y > 0.$$

6. An urn A contains six tickets numbered 1 to 6 and another urn B contains six tickets numbered 1, 2, 3, 0, 0, 0. Two tickets are drawn at random, one from each urn. Let X and Y be the random variables denoting the numbers of the tickets drawn from urn A and urn B respectively. Find the expected value of XY.

7. Show that the regression curve of Y on X for the two-dimensional distribution with joint probability density function given by

$$f(x, y)=2-x-y, \qquad 0<x<1, 0<y<1$$
$$=0, \qquad \text{elsewhere.}$$

is a hyperbola.

[*Hint* : $f_X(x)=\int_0^1 (2-x-y)\,dy=\frac{3}{2}-x,\ 0<x<1.$

$$\therefore \quad m_Y(x)=\frac{\int_0^1 y(2-x-y)\,dy}{\frac{3}{2}-x}=\frac{4-3x}{3(3-2x)}.$$

Hence, the equation of the regression curve of Y on X is

$$y=\frac{4-3x}{3(3-2x)}$$

or, $\quad 3x+9y-6xy=4,$

which is a hyperbola.]

8. If X, Y are standardised random variables, and

$$\rho\,(aX+bY, bX+aY)=\frac{1+2ab}{a^2+b^2}, \text{ then find } \rho(X, Y).$$

9. Let the marginal probability density function of X be given by

$$f_X(x)=1, \qquad -\tfrac{1}{2}<x<\tfrac{1}{2},$$

and let the conditional probability density function of Y on X be given by

$$f_Y(y \mid x)=1, \text{ if } x<y<x+1, -\tfrac{1}{2}<x<0$$
$$=1, \text{ if } -x<y<1-x, 0<x<\tfrac{1}{2}$$
$$=0, \text{ elsewhere.}$$

Show that X, Y are uncorrelated.

[*Hint* : The joint probability density function $f(x, y)$ is given by

$$\frac{f(x, y)}{f_X(x)}=f_Y(y \mid x),$$

so here $f(x, y)=1, \quad \text{if } x<y<x+1, -\frac{1}{2}<x<0$
$$=1, \quad \text{if } -x<y<1-x, 0<x<\tfrac{1}{2}$$
$$=0, \quad \text{elsewhere.}$$

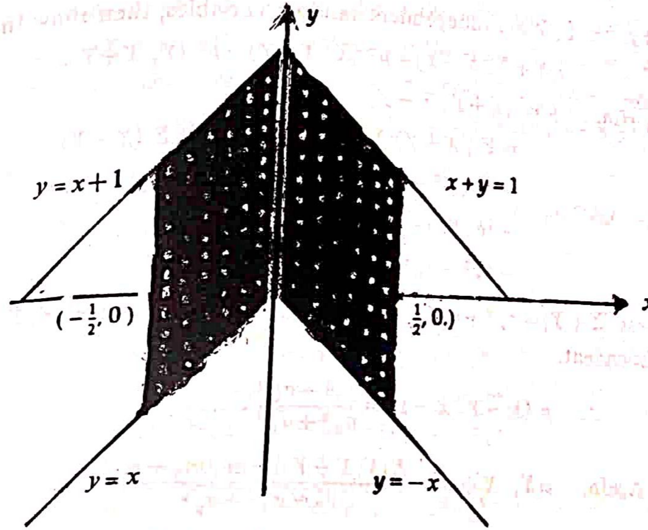


Fig. 8·10.4

$$E(X)=\int_{-\frac{1}{2}}^{\frac{1}{2}} x\,dx=0,$$

$$E(XY)=\iint_D xy\,f(x, y)\,dx\,dy,$$

where D is the shaded region shown in Fig. 8.10.4

$$=\int_{-\frac{1}{2}}^{0}\left(\int_{+x}^{x+1} xy\,dy\right)dx+\int_{0}^{\frac{1}{2}}\left(\int_{-x}^{1-x} xy\,dy\right)dx$$

$$=0. \,]$$

10. If X_1, X_2, X_3 be pairwise uncorrelated random variables, each having the same standard deviation, then find the correlation coefficient between X_1+X_2 and X_2+X_3.

11. Let X, Y, Z be three random variables each with variance σ^2 and the correlation coefficient between any two of them be a. If $U=\frac{1}{3}(X+Y+Z)$, then show that var $U=\frac{1}{3}(1+2a)\sigma^2$. Deduce that $a \geqslant -\frac{1}{2}$.

12. If X, Y are independent random variables, then show that

$$\rho(X+Y, X-Y)=\rho^2(X, X+Y)-\rho^2(Y, X+Y).$$

[*Hint* : $\text{cov}(X+Y, X-Y)$

$$=E\{(X+Y)(X-Y)\}-E(X+Y)E(X-Y)$$
$$=E(X^2-Y^2)-(m_x+m_y)(m_x-m_y)$$
$$=\{E(X^2)-m_x^2\}-\{E(Y^2)-m_y^2\}$$
$$=\sigma_x^2-\sigma_y^2.$$

var $(X+Y)=\sigma_x^2+\sigma_y^2$, var $(X-Y)=\sigma_x^2+\sigma_y^2$, since X, Y are independent.

$$\therefore \quad \rho(X+Y, X-Y)=\frac{\sigma_x^2-\sigma_y^2}{\sigma_x^2+\sigma_y^2}.$$

Again, $$\rho(X, X+Y)=\frac{E\{X(X+Y)\}-m_x(m_x+m_y)}{\sigma_x\sqrt{\sigma_x^2+\sigma_y^2}}$$

$$=\frac{\{E(X^2)-m_x^2\}+E(X)E(Y)-m_xm_y}{\sigma_x\sqrt{\sigma_x^2+\sigma_y^2}}$$

$$=\frac{\sigma_x}{\sqrt{\sigma_x^2+\sigma_y^2}}.$$

Similarly, $$\rho(Y, X+Y)=\frac{\sigma_y}{\sqrt{\sigma_x^2+\sigma_y^2}}.\ ]$$

13. Let $U=aX+bY$ and $V=bX-aY$. If $E(X)=E(Y)=0$ and if $\rho(X,Y)=\rho$, $\rho(U,V)=0$, then show that

(i) var U var $V=(a^2+b^2)^2(\text{var } X)(\text{var } Y)(1-\rho^2)$,

(ii) $ab(\text{var } X-\text{var } Y)=\rho\sigma_x\sigma_y(a^2-b^2)$,

(iii) var U+var $V=(a^2+b^2)(\text{var } X+\text{var } Y)$.

14. The random variables X, Y are normally correlated with correlation coefficient ρ. Show that U, V defined by

$$U=\frac{X}{\sigma_x}+\frac{Y}{\sigma_y} \text{ and } V=\frac{X}{\sigma_x}-\frac{Y}{\sigma_y}$$

are independent normal variates with variances $2(1+\rho)$ and $2(1-\rho)$ respectively. [C. H. (Math.) '64]

15. X, Y are correlated with correlation coefficient ρ. Show that, if a, b are constants, then the correlation coefficient between aX and bY is equal to ρ if the signs of a, b are alike and to $-\rho$ if they are different. Also show that, if the constants a, b, c are positive, then the correlation coefficient between $aX+bY$ and cY is

$$\frac{a\rho\sigma_x+b\sigma_y}{\sqrt{a^2\sigma_x^2+b^2\sigma_y^2+2ab\rho\sigma_x\sigma_y}}.$$ [C. H. (Math.) '63]

[*Hint* : $\rho(aX+bY, cY)$

$$=\frac{E\{(aX+bY)cY\}-(am_x+bm_y)cm_y}{\sqrt{a^2\sigma_x^2+b^2\sigma_y^2+2ab\rho\sigma_x\sigma_y}\sqrt{c^2\sigma_y^2}}$$

$$=\frac{ac\,E(XY)+bc\,E(Y^2)-ac\,m_xm_y-bc\,m_y^2}{\sqrt{a^2\sigma_x^2+b^2\sigma_y^2+2ab\rho\sigma_x\sigma_y}\sqrt{c^2\sigma_y^2}}$$

$$=\frac{ac\,\text{cov}(X,Y)+bc\,\sigma_y^2}{\sqrt{a^2\sigma_x^2+b^2\sigma_y^2+2ab\rho\sigma_x\sigma_y}\sqrt{c^2\sigma_y^2}}$$

$$=\frac{ac\,\rho\sigma_x\sigma_y+bc\,\sigma_y^2}{\sqrt{a^2\sigma_x^2+b^2\sigma_y^2+2ab\rho\sigma_x\sigma_y}\sqrt{c^2\sigma_y^2}}$$

$$=\frac{a\rho\sigma_x+b\sigma_y}{\sqrt{a^2\sigma_x^2+b^2\sigma_y^2+2ab\rho\sigma_x\sigma_y}}.\ ]$$

16. (X, Y) has bivariate normal distribution with parameters $m_x, m_y, \sigma_x, \sigma_y, \rho$. Show that

$$U=\frac{X-m_x}{\sigma_x},\ V=(1-\rho^2)^{-\frac{1}{2}}\left(\frac{Y-m_y}{\sigma_y}-\rho\,\frac{X-m_x}{\sigma_x}\right)$$

are independent. Find k if $\rho(U,V)=0$, when $U=X+kY$ and $V=X+\frac{\sigma_x}{\sigma_y}Y$. [C. H. (Math.) '66, '70]

[*Hint* : 2nd part.

$$\text{cov}(U,V)=E\left\{\left(X+kY\right)\left(X+\frac{\sigma_x}{\sigma_y}Y\right)\right\}-E(X+kY)E\left(X+\frac{\sigma_x}{\sigma_y}Y\right)$$

$$=E(X^2)+k\frac{\sigma_x}{\sigma_y}E(Y^2)+\left(k+\frac{\sigma_x}{\sigma_y}\right)E(XY)$$

$$-(m_x+km_y)\left(m_x+\frac{\sigma_x}{\sigma_y}m_y\right)$$

$$=E(X^2)-m_x^2+k\frac{\sigma_x}{\sigma_y}\{E(Y^2)-m_y^2\}$$

$$+k\{E(XY)-m_xm_y\}$$

$$+\frac{\sigma}{\sigma_y}\{E(XY)-m_xm_y\}$$

$$=\sigma_x^2+k\,\sigma_x\sigma_y+k\,\sigma_x\sigma_y\,\rho+\rho\sigma_x^2$$

$$=\sigma_x^2(1+\rho)+k\,\sigma_x\sigma_y\,(1+\rho)$$

$$=\sigma_x(1+\rho)(\sigma_x+k\,\sigma_y).$$

$\rho(U,V)=0$ gives

$$k=-\frac{\sigma_x}{\sigma_y}\quad(\because\ \sigma_x>0\text{ and here }-1<\rho<1).]$$

17. The least square regression lines of Y on X and of X on Y are respectively $x+3y=0$, $3x+2y=0$. If $\sigma_x=1$, then find the least square regression line of V on U where $U=X+Y$, $V=X-Y$.

[*C. H. (Math.)* '72]

18. For a certain distribution $y=12x$, $x=0{\cdot}6y$ are the regression lines. Compute $\rho(X,Y)$ and $\frac{\sigma_x}{\sigma_y}$.

Also compute $\rho(X,Z)$ if $Z=Y-X$. [*C. H. (Math.)* '82]

19. If θ be the acute angle between the two regression lines of a bivariate distribution, then prove that $\sin\theta\leqslant\frac{1-\rho^2}{1+\rho^2}$, where ρ is the correlation coefficient between the corresponding random variables.

[*Hint* : We have

$$\tan\theta=\left|\frac{\rho\frac{\sigma_y}{\sigma_x}-\frac{1}{\rho}\frac{\sigma_y}{\sigma_x}}{1+\rho\frac{\sigma_y}{\sigma_x}\cdot\frac{1}{\rho}\frac{\sigma_y}{\sigma_x}}\right|=\left|\frac{(\rho^2-1)\frac{\sigma_y}{\sigma_x}\cdot\frac{1}{\rho}}{\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2}}\right|$$

$$\tan\theta=\frac{1-\rho^2}{\rho}\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\quad(\because\ \rho^2<1)$$

$$=\frac{1-\rho^2}{2\rho}\,\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}.$$

Now $\dfrac{\sigma_x^2+\sigma_y^2}{2}\geqslant\sigma_x\sigma_y.$

$\therefore\ \tan\theta\leqslant\dfrac{1-\rho^2}{2\rho}$ (θ is a positive acute angle).

$\therefore\ \cot^2\theta\geqslant\dfrac{4\rho^2}{(1-\rho^2)^2}$ or, $\operatorname{cosec}^2\theta\geqslant\dfrac{(1+\rho^2)^2}{(1-\rho^2)^2}.$

$\therefore\ \operatorname{cosec}\theta\geqslant\dfrac{1+\rho^2}{1-\rho^2}$ or, $\sin\theta\leqslant\dfrac{1-\rho^2}{1+\rho^2}.$]

20. If the random variables $X_1, X_2, \cdots, X_{2n}$ all have the same variance σ^2 and the correlation coefficient between each pair of random variables X_i, X_j $(i\neq j)$ is ρ, then show that the correlation coefficient between $\sum_{i=1}^{n}X_i$ and $\sum_{i=n+1}^{2n}X_i$ is $\dfrac{n\rho}{1+(n-1)\rho}$.

[*Hint* : Let $U=\sum_{i=1}^{n}X_i$, $V=\sum_{i=n+1}^{2n}X_i$.

$$\sigma_u^2=n\sigma^2+2\rho\,.\,{}^nc_2\sigma^2$$

$$=n\sigma^2+n(n-1)\,\rho\sigma^2,$$

$$\sigma_v^2=n\sigma^2+n\,(n-1)\,\rho\sigma^2.$$

$$\text{cov}(U,V)=E\left(\sum_{i=1}^{n}X_i\,.\sum_{i=n+1}^{2n}X_i\right)$$

$$-(m_1+m_2+\cdots+m_n)(m_{n+1}+\cdots+m_{2n}),$$

where $E(X_i)=m_i$ for $i=1, 2, \ldots, 2n$.

Now $\dfrac{E(X_iX_j)-m_im_j}{\sigma^2}=\rho$, for $i\neq j$.

$\therefore\ E(X_iX_j)=\rho\sigma^2+m_im_j.$

$$\therefore\ E\left(\sum_{i=1}^{n}X_i\,.\sum_{i=n+1}^{2n}X_i\right)$$

$$=n^2\rho\sigma^2+(m_1+\cdots+m_n)(m_{n+1}+\cdots+m_{2n}).$$

$\therefore \quad \text{cov}(U, V) = n^2 \rho \sigma^2.$

$$\therefore \quad \rho(U, V) = \frac{n^2 \rho \sigma^2}{n\sigma^2 + \rho\sigma^2 n(n-1)}$$

$$= \frac{n\rho}{1+\rho(n-1)} \cdot]$$

21. The random variables X and Y are normally correlated and U, V are defined by

$$U = X\cos\theta + Y\sin\theta,$$

$$V = Y\cos\theta - X\sin\theta.$$

Show that U, V will be uncorrelated if

$$\tan 2\theta = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}.$$

Show further that if U, V are uncorrelated, then

$$\sigma_u\sigma_v = \sigma_x\sigma_y\sqrt{1-\rho^2}, \quad \sigma_u^2 + \sigma_v^2 = \sigma_x^2 + \sigma_y^2. \quad \text{[C. H. (Math.) '65]}$$

22. If X and Y are independent, X is normal (m_1, σ_1) and Y is normal (m_2, σ_2), and a, b are real constants, then find the characteristic function of $Z = aX + bY$ and the distribution of Z.

[C. H. (Math.) '65, '67, '83]

[*Hint* : In Theorem 8.5.7, take $n=2$, $a_1 = a$, $a_2 = b$, $X_1 = X$, $X_2 = Y$.]

23. Find the means, standard deviations and correlation coefficient of a bivariate normal distribution with parameters m_x, m_y, σ_x, σ_y, ρ. Show that the variates are independent if the correlation coefficient vanishes. [C. H. (Math.) '65]

[*Hint* : $f(x, y) = \dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times$

$$e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2 - \frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \left(\frac{y-m_y}{\sigma_y}\right)^2\right\}}$$

$$-\infty < x < \infty, \ -\infty < y < \infty.$$

The marginal density function $f_X(x)$ of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} \frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}} \times$$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y-m_y}{\sigma_y} - \rho\frac{x-m_x}{\sigma_x}\right)^2 + (1-\rho^2)\left(\frac{x-m_x}{\sigma_x}\right)^2\right\}} dy$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \times$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y - \left\{m_y - \rho\frac{\sigma_y}{\sigma_x}(x-m_x)\right\}\right]^2} dy$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-m)^2}{2\sigma^2}} dy,$$

where $\sigma = \sigma_y\sqrt{1-\rho^2}$, $m = m_y - \rho\dfrac{\sigma_y}{\sigma_x}(x-m_x)$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}},$$ since the integrand is the probability density function of a normal (m, σ) distribution.

$$\therefore \quad f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}, \ -\infty < x < \infty.$$

Similarly $f_Y(y) = \dfrac{1}{\sqrt{2\pi}\,\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}, \ -\infty < y < \infty.$

This shows that X is normal (m_x, σ_x) and Y is normal (m_y, σ_y). Hence the means of the bivariate distribution are m_x and m_y, and σ_x, σ_y are their standard deviations.

Again $\text{cov}(X, Y) = \dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_x)(y-m_y)$

$$e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-m_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \left(\frac{y-m_y}{\sigma_y}\right)^2\right\}} dx\, dy$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_x)(y-m_y)$$

$$e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-m_y}{\sigma_y}-\rho\,\frac{x-m_x}{\sigma_x}\right)^2}\; e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}\,dx\,dy.$$

Let $\xi=\dfrac{1}{\sqrt{1-\rho^2}}\left(\dfrac{y-m_y}{\sigma_y}-\rho\,\dfrac{x-m_x}{\sigma_x}\right),\ \eta=\dfrac{x-m_x}{\sigma_x}.$

Then $\dfrac{\partial(x,y)}{\partial(\xi,\eta)}=\sigma_x\sigma_y\sqrt{1-\rho^2}.$

$$\therefore\ \operatorname{cov}(X,Y)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_x\sigma_y\sqrt{1-\rho^2}\,\xi\eta+\rho\sigma_x\sigma_y\,\eta^2)\,e^{-\frac{\xi^2+\eta^2}{2}}\,d\xi\,d\eta$$

$$=\frac{\sigma_x\sigma_y\sqrt{1-\rho^2}}{2\pi}\int_{-\infty}^{\infty}\xi e^{-\frac{\xi^2}{2}}\,d\xi\int_{-\infty}^{\infty}\eta e^{-\frac{\eta^2}{2}}\,d\eta$$

$$+\rho\sigma_x\sigma_y\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{\xi^2}{2}}\,d\xi\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{\eta^2}{2}}\,d\eta$$

$$=\rho\,\sigma_x\sigma_y,$$

the first two integrals being zero and the integrands in the second two integrals being the density functions of normal (0, 1) variate.

$$\therefore\ \rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_x\sigma_y}=\rho.\ ]$$

24. Let the joint probability density function of X and Y be given by

$$f(x,y)=x^2+\frac{xy}{3},\ \text{if}\ 0<x<1,\ 0<y<2$$
$$=0\quad,\ \text{elsewhere.}$$

Find the least square regression lines of the joint distribution of X and Y.

25. Let the density function of (X, Y) be given by

$$f(x,y)=\tfrac{1}{4}\{1+xy(x^2-y^2)\},\ \text{if}\ |x|\leqslant 1,\ |y|\leqslant 1$$
$$=0\qquad\qquad,\ \text{elsewhere.}$$

Find the covariance of X and Y.

26. Let the joint distribution of X and Y be given by the probability density function $f(x, y)$ where

$$f(x,y)=y(1+x)^{-4}\,e^{-y(1+x)^{-1}},\ \text{if}\ x,y\geqslant 0$$
$$=0\qquad,\ \text{elsewhere.}$$

Find $E(Y\mid X=x)$.

27. Let the joint probability density function of X and Y be given by

$$f(x,y)=x+y,\ \text{if}\ 0<x<1, 0<y<1$$
$$=0\quad,\ \text{elsewhere.}$$

Find $\rho(X, Y)$.

28. The joint distribution of X and Y is uniform over the square with corners at the points $(0, 1), (1, 0), (-1, 0), (0, -1)$ in the xy-plane.

Show that X, Y are uncorrelated.

[*Hint*: $f(x,y)=\frac{1}{2}$, if $(x,y)\in D$
$=0$, elsewhere,

where D is the region interior to the given square.

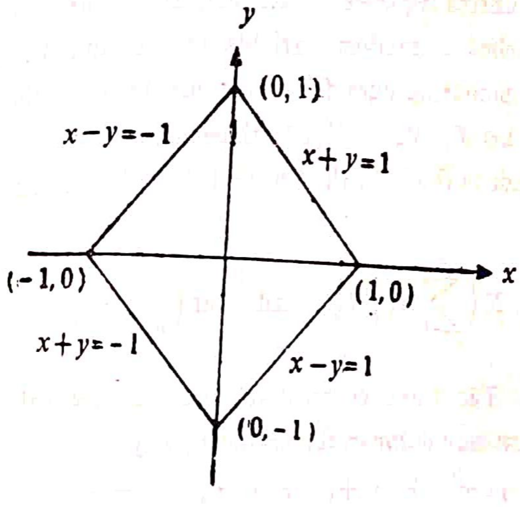


Fig. 8.10.5

Here $f_X(x)=1-x,\ 0<x<1$
$=1+x,\ -1<x<0,$
$f_Y(y)=1-y,\ 0<y<1$
$=1+y,\ -1<y<0.$

$$m_x=\int_{-1}^{1} x\,f_X(x)\,dx=0,$$

$$m_y=\int_{-1}^{1} y\,f_Y(y)\,dy=0.$$

$$\sigma_x{}^2=E(X^2)-0=\int_{-1}^{1} x^2 f_X(x)\,dx=\tfrac{1}{6},$$

$$\sigma_y{}^2=\tfrac{1}{6}.$$

$$\text{cov}\,(X, Y)=E(XY)-0$$

$$=\iint_D \frac{xy}{2}\,dx\,dy$$

$$=\frac{1}{2}\int_0^1 x\left(\int_{x-1}^{1-x} y\,dy\right)dx+\tfrac{1}{2}\int_{-1}^{0} x\left(\int_{-x-1}^{x+1} y\,dy\right)dx$$

$$=0.\ ]$$

29. A die is thrown 12 times. After each throw a $+$sign is recorded for 4, 5 or 6 and a $-$sign is recorded for 1, 2 or 3, the signs forming an ordered sequence. Each sign, except the first and the last, is attached a random variable that assumes the value 1 if both the neighbouring signs differ from one between them and 0 otherwise. Let $X_1, X_2, \ldots, X_{10}$ be these random variables, where X_i corresponds to the $(i+1)$th sign $(i=1, 2, \ldots, 10)$ in the sequence. Show that

$$E\left(\sum_{i=1}^{10} X_i\right)=\tfrac{5}{2} \quad\text{and}\quad \text{var}\left(\sum_{i=1}^{10} X_i\right)=3.$$

[*Hint*: The three consecutive signs namely i th, $(i+1)$th, $(i+2)$th signs may occur in the following ways :

$+\,+\,-$, $\;+\,-\,+$, $\;+\,+\,+$, $\;-\,-\,-$,
$-\,+\,-$, $\;-\,-\,+$, $\;+\,-\,-$, $\;-\,+\,+$.

Then $X_i=1$ for each of $+\,-\,+$, $-\,+\,-$

$=0$ otherwise.

Now $P(X_i=1)=\frac{3}{6}\cdot\frac{3}{6}\cdot\frac{3}{6}+\frac{3}{6}\cdot\frac{3}{6}\cdot\frac{3}{6}=\frac{1}{4}$,
$P(X_i=0)=\frac{6}{8}=\frac{3}{4}$.

Ex. VIII

$\therefore\ E(X_i)=1\cdot\frac{1}{4}+0\cdot\frac{3}{4}=\frac{1}{4}$ for $i=1, 2, \ldots, 10.$

$$E\left(\sum_{i=1}^{10} X_i\right)=\tfrac{10}{4}=\tfrac{5}{2}.$$

$\text{Var}\ X_i=1^2\cdot\frac{1}{4}-(\frac{1}{4})^2=\frac{3}{16},\ i=1, 2, \ldots, 10.$

Now we observe that X_i, X_j are independent if $|i-j|>2$, for example X_1, X_3 are independent, X_4, X_8 are independent etc. So, $\text{cov}\,(X_i, X_j)=0$ if $|i-j|>2$. Now for any two random variables X_i, X_j $(i\neq j)$ for which $|i-j|=1$, we can take $i=j+1$. Now

$$E(X_i X_{i+1})=1\,.\,1,\quad P(X_i=1, X_{i+1}=1).$$

Now $(X_i=1,\ X_{i+1}=1)$ occurs if i th, $(i+1)$ th, $(i+2)$th, $(i+3)$th signs occur as '$+\,-\,+\,-$' or '$-\,+\,-\,+$' and so

$$P(X_i=1, X_{i+1}=1)=\tfrac{1}{16}+\tfrac{1}{16}=\tfrac{1}{8}.$$

So, $\text{cov}\,(X_i, X_{i+1})=E(X_i X_{i+1})-\frac{1}{16}$
$=1\,.\,1\,.\,\frac{1}{8}-\frac{1}{16}=\frac{1}{16}$ for $i=1, 2, \ldots, 9.$

Hence, $\text{var}\left(\sum_{i=1}^{10} X_i\right)$

$$=\sum_{i=1}^{10}\text{var}\,X_i+2\sum_{i<j}\text{cov}\,(X_i, X_j)$$

$$=10\times\tfrac{3}{16}+2\,.\,9\,.\,\tfrac{1}{16}$$

$$=3.\ ]$$

30. Random variables X and Y have zero means and non-zero variances $\sigma_x{}^2$, $\sigma_y{}^2$. If $Z=Y-X$, find $\sigma_z{}^2$ and $\rho(X, Z)$ in terms of σ_x, σ_y. [C. H. (Math.) '82]

31. Find σ_x, σ_y and $\rho(X, Y)$ for the bivariate distribution with density function

$$f(x, y)=C\ e^{-\frac{1}{4}(x^2-2xy+4y^2)},\ -\infty<x<\infty,\ -\infty<y<\infty.$$

[C. H. (Math.) '71]

32. If U and V are independent random variables and X, Y are defined by

$$X = U \cos\theta + V \sin\theta,$$
$$Y = V \cos\theta - U \sin\theta,$$

then show that the correlation coefficient between X and Y is given by

$$\frac{(\sigma_v^2 - \sigma_u^2)\sin 2\theta}{\sqrt{(\sigma_v^2 - \sigma_u^2)^2 \sin^2 2\theta + 4\sigma_u^2\sigma_v^2}}.$$ [*C. H. (Math.) '66, 69*]

[*Hint* : See Illustrative Example 24.]

33. If $\rho(X, Y)$ denotes the coefficient of correlation between X and Y, establish that $\rho(aX+b, cY+d) = \rho(X, Y)$, where a, b, c, d are positive real constants. [*C. H. (Math.) '72*]

[*Hint* : See cor. 1 of Theorem 8.2.3.]

34. Define the joint characteristic function of the random variables X and Y. State the condition of independence of X and Y in terms of characteristic functions. If X, Y are independent and both normal (0, 1), find the characteristic function of $X+Y$. [*C. H. (Math.) '85*]

[*Hint* : See 8.5.1. Theorem 8.5.1 and Theorem 8.5.3 for $n=2$. Also see Theorem 8.5.6 for $n=2$.]

35. (*a*) Prove that the sum of two independent Poisson variates having parameters μ_1, μ_2 is a Poisson variate with parameter $\mu_1+\mu_2$. [*C. H. (Math.) '64, '66, '88, '92*]

(*b*) Prove that the sum of two binomially distributed independent random variables with parameters (n_1, p), (n_2, p) respectively is a binomially distributed random variable with parameters (n_1+n_2, p). [*C. H. (Math.) '66, '67*]

[*Hint* : Both (*a*), (*b*) have been proved in Chapter VI. Also (*a*), (*b*) can be proved with the help of characteristic function— for (*a*) take $n=2$ in Theorem 8.5.5 and for (*b*) take $n=2$ in Theorem 8.5.4.]

36. Find the equation of the regression curves of Y on X and X on Y for the bivariate normal distribution of (X, Y) with parameters $m_x, m_y, \sigma_x, \sigma_y, \rho$. Find the acute angle between them. If $\rho=0$, show that the random variables are independent.

[*C. H. (Math.) '63, '65, '66, '66 (old), '69, '69 (old)*]

[*Hint* : See Theorem 8.7.4. The required acute angle θ is the acute angle between the straight lines represented by (8.7.3) and (8.7.4). For the second part see Theorem 8.3.2.]

37. Find the conditional expectation of X given $Y=y$ for the bivariate normal distribution with probability density function given by

$$f(x, y) = \frac{1}{2\pi\,\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left(-\frac{1}{2(1-\rho^2)}\left\{\frac{x^2}{\sigma_x^2} - 2\rho\frac{xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right\}\right)$$

[*C. H. (Math.) '93*]

[*Hint* : In Theorem 8.7.4 take $m_x=0$, $m_y=0$ and show that

$$E(X \mid Y=y) = \rho\,\frac{\sigma_x}{\sigma_y}\,y. \;]$$

38. What do you understand when cov (X, Y) is stated to be zero ? Does it imply that X and Y are independent random variables ? Give reasons for your answer. [*C. H. (Econ.) '83*]

[*Hint* : See 'Significance of the correlation coefficient' in 8.8 and 'Note' after Theorem 8.2.5.]

39. Show that the covariance of two independent random variables is equal to zero. [*C. H. (Econ.) '81*]

[*Hint* : See Theorem 8.2.5.]

40. A biased coin is tossed indefinitely. Let $p(0 < p < 1)$ be the probability of success (heads). Let Y_1 denote the length of the first run (*i.e.*, the number of consecutive successes or failures starting from the first toss) and Y_2 the length of the second run. Show that

$$E(Y_1) = \frac{p}{q} + \frac{q}{p} \quad \text{and} \quad E(Y_2) = 2,$$

where $q = 1-p$.

[*Hint* : For *first part*, see Illustrative Example 32.

Second part : The event $(Y_2 = r)$ can be expressed as union of pairwise mutually exclusive events, as follows :

$$(Y_2 = r) = \sum_{x=1}^{\infty} A_x + \sum_{x=1}^{\infty} B_x,$$

where A_x denotes the event 'x successes in the first x tosses and r failures in the next r tosses and success in the $(x+r+1)$th toss' and

B_x denotes the event 'x failures in first x tosses and r successes in the next r tosses and failure in the $(x+r+1)$th toss'.

Then $P(A_x)=p^x q^r p$, $P(B_x)=q^x p^r q$.

So $$P(Y_2=r)=\sum_{x=1}^{\infty} p^{x+1}q^r+\sum_{x=1}^{\infty} q^{x+1}p^r$$
$$=p^2q^r(1-p)^{-1}+q^2p^r(1-q)^{-1}$$
$$=p^2q^{r-1}+q^2p^{r-1}.$$

$$\therefore \quad E(Y_2=r)=\sum_{r=1}^{\infty} r\,(p^2q^{r-1}+q^2p^{r-1})$$
$$=p^2\sum_{r=1}^{\infty} rq^{r-1}+q^2\sum_{r=1}^{\infty} rp^{r-1}$$
$$=p^2(1-q)^{-2}+q^2(1-p)^{-2}$$
$$=\frac{p^2}{p^2}+\frac{q^2}{q^2}$$
$$=2.\ ]$$

41. n tickets numbered 1, 2,……, n are put at random on n places marked 1, 2,……, n. Let the random variable X_i be such that $X_i=1$ if the ticket numbered 'i' is put in the place marked 'i' (*i.e.*, a match) and $X_i=0$ otherwise, for $i=1, 2,……, n$. Show that

$$\text{var}\,(X_1+X_2+\cdots+X_n)=1.$$

42. If X and Y are independent normal variates each with mean 0 and unit variance, then show that

$$E\,\{\max\,(X, Y)\}=\frac{1}{\sqrt{\pi}}.$$

43. Show that X and $Y-\rho\dfrac{\sigma_y}{\sigma_x}X$ are uncorrelated random variables.

[*Hint* : $E\left\{X\left(Y-\rho\dfrac{\sigma_y}{\sigma_x}X\right)\right\}=E\,(XY)-\rho\dfrac{\sigma_y}{\sigma_x}E(X^2)$,

$E\left(Y-\rho\dfrac{\sigma_y}{\sigma_x}X\right)=m_y-\rho\dfrac{\sigma_y}{\sigma_x}m_x$ etc.]

44. The joint probability density function of X and Y is given by

$$f(x, y)=\frac{9(1+x+y)}{2(1+x)^4(1+y)^4},\ 0<x<\infty,\ 0<y<\infty.$$

Show that $m_Y(x)=\dfrac{x+3}{2x+3}$.

45. Show that the correlation ratio of Y on X is equal to the absolute value of correlation coefficient between X and Y if (X, Y) has a bivariate normal distribution.

[*Hint* : Here $m_Y(x)=m_y+\rho\dfrac{\sigma_y}{\sigma_x}(x-m_x)$ by (8.7.3).

So $m_Y(X)=m_y+\rho\dfrac{\sigma_y}{\sigma_x}(X-m_x)$.

Then the correlation ratio of Y on X is

$$\rho\,\{m_Y(X),\ Y\}=\rho\left\{m_y+\rho\frac{\sigma_y}{\sigma_x}(X-m_x),\ Y\right\}.$$

Now $E\left\{m_y+\rho\dfrac{\sigma_y}{\sigma_x}(X-m_x)\right\}=m_y$ and $E(Y)=m_y$.

Then $$\text{cov}\left\{m_y+\rho\frac{\sigma_y}{\sigma_x}(X-m_x),\ Y\right\}$$
$$=E\left[\left\{m_y+\rho\frac{\sigma_y}{\sigma_x}(X-m_x)\right\}Y\right]-m_y^2$$
$$=m_y^2+\rho\frac{\sigma_y}{\sigma_x}E\,\{(X-m_x)\,Y\}-m_y^2$$
$$=\rho\frac{\sigma_y}{\sigma_x}\{E(XY)-m_xm_y\}$$
$$=\rho\frac{\sigma_y}{\sigma_x}\cdot\rho\sigma_x\sigma_y=\rho^2\sigma_y^2.$$

Also, $$\text{var}\left\{m_y+\rho\frac{\sigma_y}{\sigma_x}(X-m_x)\right\}$$
$$=\rho^2\frac{\sigma_y^2}{\sigma_x^2}\sigma_x^2=\rho^2\sigma_y^2.$$

$$\therefore \quad \rho\left\{m_y+\rho\frac{\sigma_y}{\sigma_x}(X-m_x),\ Y\right\}$$
$$=\frac{\rho^2\sigma_y^2}{|\rho|\,\sigma_y\cdot\sigma_y}=|\rho|.\ ]$$

46. Show that the regression curves for the means coincide with the corresponding least square regression lines for the bivariate distribution of (X, Y) with the probability density function given by

$$f(x, y) = 6(1-x-y), \quad \text{if} \quad x > 0, y > 0, x+y < 1$$
$$= 0 \qquad , \quad \text{elsewhere.}$$

47. Show that for a bivariate normal distribution, the moments μ_{rs} obey the recurrence relation

$$\mu_{rs} = \rho(r+s-1)\mu_{r-1,\, s-1} + (r-1)(s-1)(1-\rho^2)\mu_{r-2,\, s-2}.$$

[*C. H. (Math.)* '67]

[*Hint*: Here we assume that $m_x = 0$, $m_y = 0$, $\sigma_x = 1$, $\sigma_y = 1$. If $M(t_1, t_2)$ be the two-dimensional moment generating function of a bivariate normal distribution with $m_x = m_y = 0$, $\sigma_x = \sigma_y = 1$, then by Theorem 8.3.1, we get

$$M(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + t_2^2 + 2\rho t_1 t_2)}.$$

Then $\dfrac{1}{M}\dfrac{\partial M}{\partial t_1} = t_1 + \rho t_2$, $\dfrac{1}{M}\dfrac{\partial M}{\partial t_2} = t_2 + \rho t_1$,

and
$$\frac{\partial^2 M}{\partial t_1 \partial t_2} = \frac{\partial M}{\partial t_1}(t_2 + \rho t_1) + \rho M$$
$$= t_2 \frac{\partial M}{\partial t_1} + \rho t_1 \frac{\partial M}{\partial t_1} + \rho M.$$

But $t_2 \dfrac{\partial M}{\partial t_1} - \rho t_2 \dfrac{\partial M}{\partial t_2} = M(1-\rho^2) t_1 t_2$.

$$\therefore \quad \frac{\partial^2 M}{\partial t_1 \partial t_2} = \rho t_1 \frac{\partial M}{\partial t_1} + \rho t_2 \frac{\partial M}{\partial t_2} + M(1-\rho^2)t_1 t_2 + \rho M.$$

Now differentiating both sides partially with respect to t_1, $(r-1)$ times (by Leibnitz theorem), we get

$$\frac{\partial^{r+1} M}{\partial t_1^r \partial t_2} = \rho\left\{t_1 \frac{\partial^r M}{\partial t_1^r} + (r-1)\frac{\partial^{r-1} M}{\partial t_1^{r-1}}\right\} + \rho t_2 \frac{\partial^r M}{\partial t_1^{r-1} \partial t_2}$$
$$+ (1-\rho^2)\, t_2 \left\{t_1 \frac{\partial^{r-1} M}{\partial t_1^{r-1}} + (r-1)\frac{\partial^{r-2} M}{\partial t_1^{r-2}}\right\} + \rho \frac{\partial^{r-1} M}{\partial t_1^{r-1}}$$

$(r > 1)$.

Again differentiating with respect to t_2, $(s-1)$ times, we get

$$\frac{\partial^{r+s} M}{\partial t_1^r \partial t_2^s} = \rho\left\{t_1 \frac{\partial^{r+s-1} M}{\partial t_1^r \partial t_2^{s-1}} + (r-1)\frac{\partial^{r+s-2} M}{\partial t_1^{r-1} \partial t_2^{s-1}}\right\}$$
$$+ \rho\left\{t_2 \frac{\partial^{r+s-1} M}{\partial t_1^{r-1} \partial t_2^s} + (s-1)\frac{\partial^{r+s-2} M}{\partial t_1^{r-1} \partial t_2^{s-1}}\right\}$$
$$+ (1-\rho^2)t_1\left\{t_2 \frac{\partial^{r+s-2} M}{\partial t_1^{r-1} \partial t_2^{s-1}} + (s-1)\frac{\partial^{r+s-3} M}{\partial t_1^{r-1} \partial t_2^{s-2}}\right\}$$
$$+ (1-\rho^2)(r-1)\left\{t_2 \frac{\partial^{r+s-3} M}{\partial t_1^{r-2} \partial t_2^{s-1}} + (s-1)\frac{\partial^{r+s-4} M}{\partial t_1^{r-2} \partial t_2^{s-2}}\right\}$$
$$+ \rho \frac{\partial^{r+s-2} M}{\partial t_1^{r-1} \partial t_2^{s-1}}.$$

Putting $t_1 = 0$, $t_2 = 0$, we get

$$\alpha_{rs} = \rho(r-1)\,\alpha_{r-1,\, s-1} + \rho(s-1)\,\alpha_{r-1,\, s-1}$$
$$+ (r-1)(1-\rho^2)(s-1)\,\alpha_{r-2,\, s-2} + \rho\alpha_{r-1,\, s-1}$$
$$= \rho(r+s-1)\,\alpha_{r-1,\, s-1} + (r-1)(s-1)(1-\rho^2)\,\alpha_{r-2,\, s-2}.$$

Here $\mu_{rs} = \alpha_{rs}$.]

48. (*a*) If $X_1, X_2, \ldots X_n$ are mutually independent identically distributed random variables and if $X_1 + X_2 + \cdots + X_n$ is a Poisson μ-variate, then show that each X_i is a Poisson $\dfrac{\mu}{n}$ variate.

(*b*) If $X_1, X_2, \ldots, X_n$ are mutually independent identically distributed random variables and if $X_1 + X_2 + \cdots + X_n$ is a binomial (r, p) variate, then show that each X_i is a binomial $\left(\dfrac{r}{n}, p\right)$ variate, provided $\dfrac{r}{n}$ is an integer.

49. The joint probability density function of X and Y is given by

$$f(x, y) = axy, \quad \text{if} \quad 0 < x < 1, 0 < y < 1,$$

where a is a constant. Find $E(\sqrt{X^2 + Y^2})$.

50. n numbers are chosen independently at random from the interval (a, b) and the numbers are arranged in ascending order. Let X_i be the random variable denoting the ith number in the above sequence, for $i = 1, 2, \ldots\ldots, n$.

Prove that $E(X_i)=a+\frac{i}{n+1}(b-a)$, for $i=1, 2, \ldots\ldots, n$.

[*Hint*: Let $Y_1, Y_2, \ldots, Y_n$ be the random variables denoting n numbers chosen at random from (a, b). Then $Y_1, Y_2, \ldots, Y_n$ have uniform continuous distribution in (a, b) and they are mutually independent. Now from the definition of $X_1, X_2, \ldots, X_n$, we find that

$$X_1=\min(Y_1, Y_2, \ldots, Y_n).$$

Then $(X_1 > x) \Leftrightarrow (Y_1 > x, Y_2 > x, \ldots, Y_n > x)$.

Similarly,

$$\begin{aligned} X_2 > x \Leftrightarrow & (Y_1 > x, Y_2 > x, \ldots, Y_n > x) \\ & +(Y_1 < x, Y_2 > x, \ldots, Y_n > x) \\ & +(Y_2 < x, Y_1 > x, Y_3 > x, \ldots, Y_n > x) \\ & +\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \\ & +(Y_n < x, Y_1 > x, Y_2 > x, \ldots, Y_{n-1} > x), \end{aligned}$$

and so on.

$$\therefore \quad P(X_1 > x)=P(Y_1 > x)\, P(Y_2 > x) \ldots P(Y_n > x)$$

$$=\left(\int_x^b \frac{dx}{b-a}\right)^n=\left(\frac{b-x}{b-a}\right)^n \text{ if } a < x < b$$

$$P(X_2 > x)=\left(\frac{b-x}{b-a}\right)^n+n\left(\frac{b-x}{b-a}\right)^{n-1}\left(\frac{x-a}{b-a}\right) \text{ if } a < x < b,$$

and so on.

Then $1-F_1(x)=\frac{(b-x)^n}{(b-a)^n}$ if $a < x < b$,

$$1-F_2(x)=\frac{(b-x)^n}{(b-a)^n}+\frac{n(b-x)^{n-1}}{(b-a)^n}(x-a) \text{ if } a < x < b$$

and so on,

where $F_1(x), F_2(x), \ldots\ldots, F_n(x)$ are the distribution functions of $X_1, X_2, \ldots\ldots, X_n$ respectively. In general we get

$$1-F_k(x)=\frac{(b-x)^n}{(b-a)^n}+\frac{n(b-x)^{n-1}}{(b-a)^n}(x-a)+{}^nC_2\frac{(b-x)^{n-2}}{(b-a)^n}(x-a)^2$$

$$+\cdots\cdots+{}^nC_{k-1}\frac{(b-x)^{n-k+1}}{(b-a)^n}(x-a)^{k-1} \text{ if } a < x < b,$$

for $k=2, 3, \ldots\ldots, n-i$,

and $F_n(x)=P(X_n < x)$

$$=P(Y_1 < x, Y_2 < x, \ldots\ldots, Y_n < x)$$

$$=\frac{(x-a)^n}{(b-a)^n} \quad \text{if } a < x < b.$$

Also we note that for each $F_k(x)$,

$$F_k(x)=0 \quad \text{if } x < a$$

$$=1 \quad \text{if } x \geqslant b.$$

As an example let us find the value of $E(X_2)$. The probability density function $f_2(x)$ of X_2 is given by

$$f_2(x)=F_2'(x)=\frac{n(b-x)^{n-1}}{(b-a)^n}-\frac{n(b-x)^{n-1}}{(b-a)^n}+\frac{n(n-1)(b-x)^{n-2}}{(b-a)^n}(x-a)$$

if $a < x < b$

So, $$E(X_2)=\int_a^b \frac{n(n-1)}{(b-a)^n}(b-x)^{n-2}(x-a)\, x\, dx$$

$$=\frac{2n(n-1)}{(b-a)^n}\int_0^{\frac{\pi}{2}}(b-a)^n \cos^{2n-3}\theta \sin^3\theta\,(a\cos^2\theta+b\sin^2\theta)\, d\theta,$$

where $x=a\cos^2\theta+b\sin^2\theta$

$$=2n(n-1)\int_0^{\frac{\pi}{2}}(a\cos^{2n-1}\theta\sin^3\theta+b\cos^{2n-3}\theta\sin^5\theta)\, d\theta$$

$$=n(n-1)\left\{a\cdot\frac{\Gamma(n)\Gamma(2)}{\Gamma(n+2)}+b\cdot\frac{\Gamma(n-1)\Gamma(3)}{\Gamma(n+2)}\right\}$$

$$=\frac{a(n-1)}{n+1}+\frac{2b}{n+1}$$

$$=a+\frac{2}{n+1}(b-a).\ ]$$

Answers

1. $E(Y \mid X=x)=\frac{x}{2}$ if $0 \leqslant x \leqslant a$, $\frac{1}{2}$, $13u+15v=0$.

2. (i) -1, (ii) 0. 6. $\frac{7}{2}$. 8. $\frac{a^2+b^2}{(a^2-b^2)^2-2ab}$.

10. $\frac{1}{2}$. 17. $5v-3u=0$. 18. $\frac{2}{\sqrt{159}}$.

24. $y-\frac{10}{9}=-\frac{10}{73}\left(x-\frac{13}{18}\right)$, $-\frac{13}{18}=-\frac{1}{32}\left(y-\frac{10}{9}\right)$.

25. 0. 26. $2(1+x)$. 27. $-\frac{1}{11}$.

30. $\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y$, $\frac{\rho\sigma_y-\sigma_x}{\sqrt{\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y}}$.

31. $\sigma_x=\frac{2\sqrt{2}}{\sqrt{3}}$, $\sigma_y=\frac{\sqrt{2}}{\sqrt{3}}$, $\rho=\frac{1}{2}$.

49. $\frac{8(2\sqrt{2}-1)}{15}$.

Answers

1. $E(Y \mid X=x)=\frac{x}{2}$ if $0 \leqslant x \leqslant a$, $\frac{1}{3}$, $13u+15v=0$.

2. (*i*) -1, (*ii*) 0. 6. $\frac{7}{2}$. 8. $\frac{a^2+b^2}{(a^2-b^2)^2-2ab}$.

10. $\frac{1}{2}$. 17. $5v-3u=0$. 18. $\frac{2}{\sqrt{159}}$.

24. $y-\frac{10}{9}=-\frac{10}{73}\left(x-\frac{13}{18}\right)$, $-\frac{13}{18}=-\frac{1}{32}\left(y-\frac{10}{9}\right)$.

25. 0. 26. $2(1+x)$. 27. $-\frac{1}{11}$.

30. $\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y$, $\frac{\rho\sigma_y-\sigma_x}{\sqrt{\sigma_x^2+\sigma_y^2-2\rho\sigma_x\sigma_y}}$.

31. $\sigma_x=\frac{2\sqrt{2}}{\sqrt{3}}$, $\sigma_y=\frac{\sqrt{2}}{\sqrt{3}}$, $\rho=\frac{1}{2}$.

49. $\frac{8(2\sqrt{2}-1)}{15}$.

CHAPTER IX

SOME IMPORTANT CONTINUOUS UNIVARIATE DISTRIBUTIONS

9.1. Normal Distribution.

Normal distribution has already been defined in chapter V and important characteristics like mean, variance, skewness, kurtosis, median, etc. of this distribution have been obtained in chapter VII. In real life situations we come across many probability distributions which are approximately normal, for example, the distributions of the random variables giving (*i*) the measured value of a physical quantity, (*ii*) the I.Q. of a given group of students, (*iii*) the increments in blood pressure of a given group of patients by the application of a given drug, etc. are approximately normal. In fact, of all the continuous distributions, the most important distribution which can be used in practice is the normal distribution. Further, from De Moivre-Laplace theorem (Chapter X) we know that binomial (n, p) distribution can be approximated by normal distribution if n is large and p fixed and Central Limit Theorem (Chapter X) also reveals that $\frac{X_1+X_2+\cdots+X_n}{n}$ is approximately $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$ if n is very large and where $X_1, X_2, \ldots, X_n$ are mutually independent and have the same distribution with mean m and standard deviation σ.

Because of this importance of normal distribution, we summarise the properties of this distribution already obtained in the previous chapters and also explain some properties in more detail in this section.

Properties of Normal Distribution :

(a) *A normal distribution is symmetrical about the mean of the distribution.*

The probability density function of a normal (m, σ) distribution is given by

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}},\quad -\infty < x < \infty,\ \sigma > 0.$$

Now $f(m+x)=\dfrac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}$, $f(m-x)=\dfrac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}$.

So $f(m+x)=f(m-x)$ for all $x \in (-\infty, \infty)$. Hence the distribution is symmetrical about m which is the mean of the normal (m, σ) distribution.

(b) *Median and mode of a normal (m, σ) distribution are both equal to the mean m (proved in chapter VII), as expected for a symmetrical distribution.*

(c) *Odd order central moments are zero. Even order central moments are given by*

$$\mu_{2k}=1\,.\,3\,.\,5 \ldots (2k-1)\,\sigma^{2k} \quad \text{for } k=1, 2, 3, \ldots$$

[For proof, see 7.3.60].

In particular $\mu_3=0$, $\mu_4=3\sigma^4$.

So $\gamma_1=\dfrac{\mu_3}{\sigma^3}=0$, $\beta_2=\dfrac{\mu_4}{\sigma^4}=3$ and $\gamma_2=\beta_2-3=0$.

So for a normal distribution, coefficient of skewness is 0, as expected for a symmetrical distribution and the coefficient of excess is also equal to zero.

(d) The moment generating function (about the origin) and the characteristic function of a normal (m, σ) distribution are respectively given by

$$M(t)=e^{mt+\frac{1}{2}\sigma^2t^2},\ t\varepsilon(-\infty, \infty),$$

$$\phi(t)=e^{imt-\frac{1}{2}\sigma^2t^2}\quad t\varepsilon(\infty, \infty).$$

(e) *The density curve* $y=\dfrac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$ *has two points of inflexion equidistant from the line $x=m$.*

For points of inflexion we have $\dfrac{d^2y}{dx^2}=0$, $\dfrac{d^3y}{dx^3}\neq 0$. Here

$$\frac{dy}{dx}=-\frac{x-m}{\sqrt{2\pi}\,\sigma^3}e^{-\frac{(x-m)^2}{2\sigma^2}},$$

$$\frac{d^2y}{dx^2}=-\frac{1}{\sqrt{2\pi}\,\sigma^3}e^{-\frac{(x-m)^2}{2\sigma^2}}+\frac{(x-m)^2}{\sqrt{2\pi}\,\sigma^5}e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Now $\dfrac{d^2y}{dx^2}=0$ gives $-1+\dfrac{(x-m)^2}{\sigma^2}=0$

or, $x=m\pm\sigma$.

It can be shown that $\dfrac{d^3y}{dx^3}\neq 0$ for $x=m\pm\sigma$.

So the density curve $y=\dfrac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$ has points of inflexion at $x=m+\sigma$, $x=m-\sigma$ which are at a distance σ from the straight line $x=m$. So we can state that the density curve of a normal (m, σ) distribution is convex upwards within $(m-\sigma, m+\sigma)$ and concave upwards outside $(m-\sigma, m+\sigma)$.

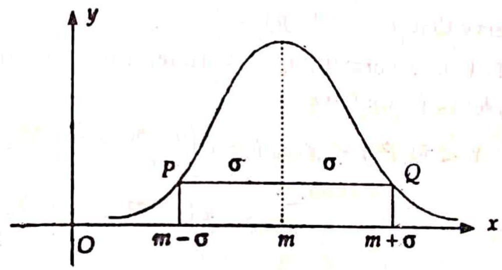


Fig. 9.1.1 Normal density curve.

Fig. 9.1.1 represents the density curve of a normal (m, σ) distribution where P, Q are the points of inflexion of the curve.

(f) *If X_1, [illegible] independent normal variates with parameters (m_1, σ_1), (m_2, σ_2), ..., (m_n, σ_n) respectively and $a_1, a_2, \ldots, a_n$ are real constants, then $a_1X_1+a_2X_2+\cdots+a_nX_n$ is normal $\left(a_1m_1+a_2m_2+\cdots+a_nm_n, \sqrt{a_1^2\sigma_1^2+a_2^2\sigma_2^2+\cdots+a_n^2\sigma_n^2}\right)$ variate* (Theorem 8.5.7).

(g) Let $U=\frac{X-m}{\sigma}$ where X is a normal (m, σ) variate. In real variables we have

$$u=\frac{x-m}{\sigma}.$$

Here $\frac{du}{dx}=\frac{1}{\sigma}>0$ for all x. So if $f_U(u)$ be the probability density function of U, we have

$$f_U(u)=\sigma \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{u^2}{2}}, \quad -\infty < u < \infty,$$

that is, $$f_U(u)=\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \quad -\infty < u < \infty,$$

which shows that U is a standard normal variate.

Writing $\phi(u)$ for $f_U(u)$ we find that the distribution function $\Phi(u)$ of U is given by

$$\Phi(u)=\int_{-\infty}^{u} \phi(v)\, dv=\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{v^2}{2}} dv, \quad -\infty < u < \infty. \qquad (9.1.1)$$

We observe that $\Phi(\infty)=1$, $\Phi(0)=\frac{1}{2}$.

Now if X be a normal (m, σ) variate, then for any two real numbers a, b $(a<b)$ we have

$$P(a < X < b), P(a < X \leqslant b)=P\left(\frac{a-m}{\sigma} < \frac{X-m}{\sigma} \leqslant \frac{b-m}{\sigma}\right)$$

$$=P\left(\frac{a-m}{\sigma} < U \leqslant \frac{b-m}{\sigma}\right)$$

$$=\Phi\left(\frac{b-m}{\sigma}\right)-\Phi\left(\frac{a-m}{\sigma}\right)=\int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} \phi(v)\, dv$$

which represents the area of the region (shaded in Fig. 9.1.2) bounded by $y=0$, $x=\frac{a-m}{\sigma}$, $x=\frac{b-m}{\sigma}$ and the standard normal density curve $y=\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

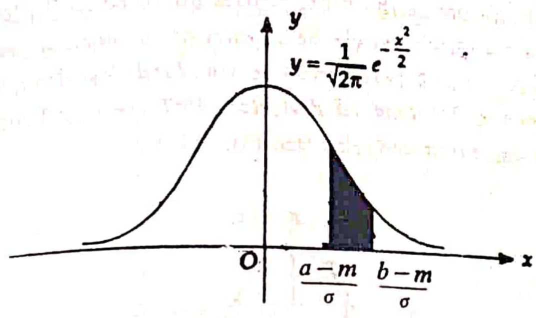


Fig. 9.1.2. $P(a < X \leqslant b)$.

The values of $\Phi(x)=\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$, for different values of x, are found from Table I given in the Appendix and using the relation

$$P(a < X < b)=P(a < X \leqslant b)=\Phi\left(\frac{b-m}{\sigma}\right)-\Phi\left(\frac{a-m}{\sigma}\right), \qquad (9.1.2)$$

we can determine the value of $P(a < X \leqslant b)$. In this connection we observe that $\Phi(-x)=1-\Phi(x)$. Then we find that

(i) $P(m-\sigma < X < m+\sigma)=P(-1 < U < 1)$
$$=\Phi(1)-\Phi(-1)=2\Phi(1)-1$$
$$=2\times 0{\cdot}8413447-1$$
$$=1{\cdot}6826894-1$$
$$\simeq 0{\cdot}68.$$

(ii) $P(m-2\sigma < X < m+2\sigma)=P(-2 < U < 2)$
$$=2\Phi(2)-1$$
$$=2\times 0{\cdot}9772499-1$$
$$=1{\cdot}9544998-1$$
$$\simeq 0{\cdot}95$$

(iii) $P(m-3\sigma < X < m+3\sigma)=P(-3 < U < 3)$
$$=2\Phi(3)-1$$
$$=2\times 0{\cdot}9986501-1$$
$$=1{\cdot}9973002-1$$
$$\simeq 0{\cdot}997.$$

From the values of the probabilities obtained in (*i*), (*ii*), (*iii*) we can state that approximeately 68% *values of a normal variate fall between* $m \pm \sigma$, *i.e., between mean* $\pm$ *standard deviation*, 95% *values between mean* $\pm$ $2\times$ *standard deviation*, 99·7% (*nearly* 100%) *between mean* $\pm$ $3\times$ *standard deviation* (see Fig. 9.1.3).

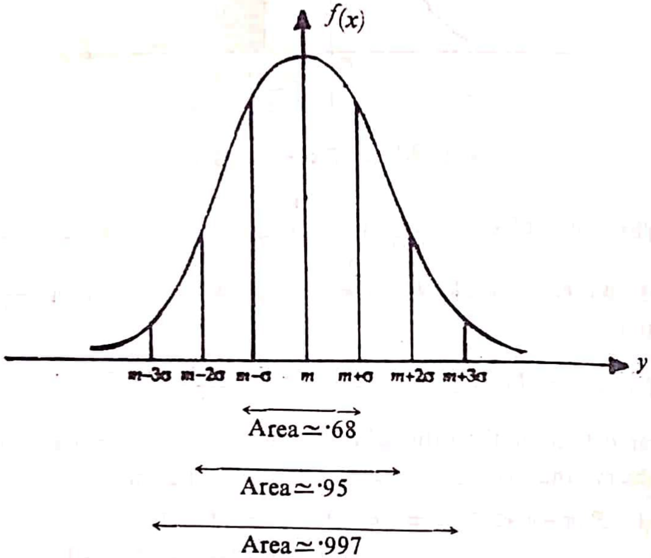


Fig. 9.1.3. Probability density curve a normal (m, σ) distribution.

9.2. Chi-square Distribution.

The chi-square distribution (also stated as χ^2-distribution) is a continuous distribution and this was first obtained by Helmert in 1875 and later by Karl Pearson in 1900. Here we shall follow the convention of denoting the corresponding random variable by χ^2 (instead of X, Y etc.) and the corresponding real vaiiable again by χ^2. So we can speak "a random variable χ^2 has a χ^2-distribution".

Now we give the formal definition of a chi-square distribution with the help of probability density function.

A χ^2-distribution is a continuous distribution whose probability density function $f(\chi^2)$ is given by

$$f(\chi^2)=\frac{e^{-\frac{\chi^2}{2}}\left(\frac{\chi^2}{2}\right)^{\frac{n}{2}-1}}{2\Gamma(\frac{n}{2})} \quad \text{if} \quad \chi^2>0$$
$$=0, \qquad \text{elsewhere}\; ; \qquad \cdots \qquad (9.2.1)$$

where n is a positive integer and n is the only parameter of this distribution. The positive integer n is called the number of *degrees of freedom* of the distribution.

If a random variable X has chi-square distribution with n degrees of freedom, we say that X is a $\chi^2(n)$ variate and the distribution is also mentioned as a $\chi^2(n)$ distribution.

THEOREM 9.2.1. *If X has χ^2-distribution with n degrees of freedom, then* $\frac{X}{2}$ is $\gamma\left(\frac{n}{2}\right)$ *variate and conversely if* $\frac{X}{2}$ *is a* $\gamma\left(\frac{n}{2}\right)$ *variate* (*n is a positive integer*), *then X has χ^2-distribution with n degrees of freedom.*

Proof: Let X be a $\chi^2(n)$ variate.

Let $Y=\frac{X}{2}$.

In real variables, we have $y=\frac{x}{2}$.

Then $\frac{dy}{dx}=\frac{1}{2}>0$ for all x. So if $f_Y(y)$ be the probability density function of Y, then

$$f_Y(y)=2\,f(\chi^2),$$

where the probability density function $f(\chi^2)$ of X is given by

$$f(\chi^2)=\frac{e^{-\frac{\chi^2}{2}}\left(\frac{\chi^2}{2}\right)^{\frac{n}{2}-1}}{2\Gamma(\frac{n}{2})} \quad \text{if} \quad \chi^2>0$$
$$=0, \qquad \text{elsewhere.}$$

Then $f_Y(y) = \dfrac{e^{-\frac{x^2}{2}}\left(\frac{x^2}{2}\right)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})}$

$= \dfrac{e^{-\frac{x}{2}}\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})}$, writing x for x^2

$= \dfrac{e^{-y}\, y^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})}$ if $0 < y < \infty$.

So the probability density function of $Y = \frac{X}{2}$ is given by

$$f_Y(y) = \frac{e^{-y}\, y^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \quad \text{if } 0 < y < \infty$$

$$= 0, \qquad \text{elsewhere ;}$$

and this shows that $Y = \frac{X}{2}$ is a $\gamma\left(\frac{n}{2}\right)$ variate.

Now let $\frac{X}{2}$ be a $\gamma\left(\frac{n}{2}\right)$ variate. Again using the transformation $Y = \frac{X}{2}$, *i.e.*, $X = 2Y$ we find that Y is a $\gamma\left(\frac{n}{2}\right)$ variate and in real variables we get $x = 2y$ which gives

$$\frac{dx}{dy} = 2 > 0 \text{ for all } y.$$

Then the probability density function $f_X(x)$ of X is given by

$f_X(x) = \frac{1}{2} f_Y(y)$, where

$$f_Y(y) = \frac{e^{-y}\, y^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)} \text{ if } 0 < y < \infty$$

$$= 0, \qquad \text{elsewhere.}$$

So $$f_X(x) = \frac{e^{-\frac{x}{2}}\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n}{2}\right)} \quad \text{if } x > 0$$

$$= 0, \qquad \text{elsewhere}$$

and this shows that X has χ^2-distribution with n degrees of freedom.

THEOREM 9.2.2. *If $X_1, X_2, \ldots, X_n$ are mutually independent standard normal variates, then $X_1^2 + X_2^2 + \cdots + X_n^2$ has χ^2-distribution with n degrees of freedom.*

Proof: We shall prove the theorem by induction. If $n = 1$, we have only one standard normal variate X_1. Here we write X for X_1. Let $Y = X^2$, where $X = X_1$ is a standard normal variate. Now the spectrum of Y is the set $\{y : 0 \leqslant y < \infty\}$.

Let $F_Y(y)$ be the distribution function of Y.

Then $F_Y(y) = P(Y \leqslant y) = P(X^2 \leqslant y) = P(-\sqrt{y} \leqslant X \leqslant \sqrt{y})$ if $y \geqslant 0$.

Then if $y \geqslant 0$, $F_Y(y) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{t^2}{2}}\, dt.$

$$\therefore \quad F_Y(y) = \sqrt{\frac{2}{\pi}} \int_0^{\sqrt{y}} e^{-\frac{t^2}{2}}\, dt \quad \text{if } y \geqslant 0. \tag{9.2.2}$$

If $y < 0$, then

$F_Y(y) = P(Y < 0) = P(X^2 < 0) = 0$, since $P(X^2 < 0) = 0$.

Let $f_Y(y)$ be the probability density function of Y.

Then $f_Y(y) = F'_Y(y) = 0$ if $y < 0$.

Again from (9.2.2), we find

$$F_Y'(y) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{2\sqrt{y}}\, e^{-\frac{y}{2}} \text{ if } y > 0$$

$$= \frac{1}{2\sqrt{\pi}}\, e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{-\frac{1}{2}} \text{ if } y > 0$$

$$= \frac{1}{2\Gamma(\frac{1}{2})}\, e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{-\frac{1}{2}} \text{ if } y > 0.$$

Thus $$F_Y'(y) = \frac{e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{-\frac{1}{2}}}{2\Gamma(\frac{1}{2})} \text{ if } y > 0$$

$$= 0, \qquad \text{if } y < 0.$$

Again defining $f_Y(0) = 0$, we find that

$$f_Y(y) = \frac{e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{\frac{1}{2}-1}}{2\Gamma(\frac{1}{2})} \text{ if } y > 0$$

$$= 0, \qquad \text{elsewhere,}$$

and this shows that Y has χ^2-distribution with 1 degree of freedom, i.e., $X^2=X_1^2$ has $\chi^2(1)$ distribution. So the required proposition is true for $n=1$.

Now let the proposition be true for any m (a positive integer) mutually independent standard normal variates $X_1, X_2, \ldots, X_m$. Then $X_1^2+X_2^2+\cdots+X_m^2$ is a $\chi^2(m)$ variate. Now let $X_1, X_2, \ldots, X_m, X_{m+1}$ be any $m+1$ mutually independent standard normal variates.

Let $Y=X_1^2+X_2^2+\cdots+X_m^2+X_{m+1}^2$.

Then we can write $Y=U+V$, where $U=X_1^2+X_2^2+\cdots+X_m^2$, which by induction hypothesis is a $\chi^2(m)$ variate and $V=X_{m+1}^2$ is a $\chi^2(1)$ variate, since the proposition is true for one random variable. Also U, V are independent. Again let $Z=V$. Then we have the transformation

$$Y=U+V, \quad Z=V.$$

In real variables $y=u+v$, $z=v$.

$$\therefore \quad \frac{\partial(y, z)}{\partial(u, v)}=\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}=1>0 \text{ for all } u, v.$$

Then if $f_{Y,Z}(y, z)$ be the joint probability density function of Y and Z, we get

$$f_{Y,Z}(y, z)=f_U(u)\, f_V(v)$$

$$=\frac{e^{-\frac{u}{2}}\left(\frac{u}{2}\right)^{\frac{m}{2}-1}}{2\Gamma\left(\frac{m}{2}\right)} \cdot \frac{e^{-\frac{v}{2}}\left(\frac{v}{2}\right)^{\frac{1}{2}-1}}{2\Gamma\left(\frac{1}{2}\right)} \quad \text{if } u>0, v>$$

$$=\frac{e^{-\frac{u+v}{2}}\left(\frac{u}{2}\right)^{\frac{m}{2}-1}\left(\frac{v}{2}\right)^{-\frac{1}{2}}}{4\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1}{2}\right)} \quad \text{if } u>0, v>0$$

$$=\frac{e^{-\frac{y}{2}}\left(\frac{y-z}{2}\right)^{\frac{m}{2}-1}\left(\frac{z}{2}\right)^{-\frac{1}{2}}}{4\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1}{2}\right)} \quad \text{if } y>z, z>0.$$

So the probability density function $f_Y(y)$ of Y is given by

$$f_Y(y)=\frac{e^{-\frac{y}{2}}}{4\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1}{2}\right)}\int_0^y\left(\frac{y-z}{2}\right)^{\frac{m}{2}-1}\left(\frac{z}{2}\right)^{-\frac{1}{2}} dz$$

$$=\frac{e^{-\frac{y}{2}}\sqrt{2}}{2\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1}{2}\right)2^{\frac{m}{2}}}\int_0^y (y-z)^{\frac{m}{2}-1} z^{\frac{1}{2}-1}\, dz$$

$$=\frac{e^{-\frac{y}{2}}y^{\frac{m}{2}-1}y^{\frac{1}{2}-1}\, y}{\sqrt{2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1}{2}\right)2^{\frac{m}{2}}}\int_0^1 (1-t)^{\frac{m}{2}-1} t^{\frac{1}{2}-1}\, dt,$$

where $z=yt$

$$=\frac{e^{-\frac{y}{2}}y^{\frac{m+1}{2}-1}}{\sqrt{2}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{1}{2}\right)2^{\frac{m}{2}}}\, B\left(\frac{1}{2}, \frac{m}{2}\right)$$

$$=\frac{e^{-\frac{y}{2}}y^{\frac{m+1}{2}-1}}{2^{\frac{m+1}{2}}\Gamma\left(\frac{m+1}{2}\right)} \quad \text{if } y>0.$$

Hence the probability density function of

$$Y=X_1^2+X_2^2+\cdots+X_m^2+X_{m+1}^2$$

is given by

$$f_Y(y)=\frac{e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{\frac{m+1}{2}-1}}{2\Gamma\left(\frac{m+1}{2}\right)} \quad \text{if } y>0$$

$$=0, \qquad \text{elsewhere},$$

and this shows that $Y=X_1^2+X_2^2+\cdots+X_{m+1}^2$ has $\chi^2(m+1)$ distribution. So it is proved that if the required proposition is true for $n=m$, it is also true for $n=m+1$. Also we have seen that it is true for $n=1$. Hence by the principle of mathematical induction, it is true for every positive integer n.

Mean, Variance, Mode, Characteristic Function of the Chi-square distribution.

Let X be a $\chi^2(n)$ variate. Then by Theorem 9.2.1, $\frac{X}{2}$ is a $\gamma\left(\frac{n}{2}\right)$ variate. Now we know that mean and variance of a $\gamma(l)$ variate are both equal to l.

So $E\left(\frac{X}{2}\right)=\frac{n}{2}$ and var $\frac{X}{2}=\frac{n}{2}$.

But $E\left(\frac{X}{2}\right)=\frac{1}{2}E(X)$ and so $E(X)=n$.

Also var $\frac{X}{2}=\frac{1}{4}$ var $X=\frac{n}{2}$. $\therefore$ var $X=2n$.

So the mean and variance of $\chi^2(n)$ distribution are respectively n and $2n$.

Now from (9.2.1) we find that if $n \leqslant 2$, then $f(\chi^2)$ decreases as χ^2 increases, *i.e.*, $f(\chi^2)$ is a monotonically decreasing function in $0<\chi^2<\infty$ and consequently $\chi^2(n)$ distribution has no mode if $n \leqslant 2$. But if $n>2$, we find that $f'(\chi^2)=0$ ($\chi^2>0$) gives

$$e^{-\frac{\chi^2}{2}}\left\{\left(\frac{n}{2}-1\right)\left(\frac{\chi^2}{2}\right)^{\frac{n}{2}-2}\cdot\frac{1}{2}-\frac{1}{2}\left(\frac{\chi^2}{2}\right)^{\frac{n}{2}-1}\right\}=0$$

or, $\left(\frac{n}{2}-1\right)\left(\frac{x}{2}\right)^{\frac{n}{2}-2}-\left(\frac{x}{2}\right)^{\frac{n}{2}-1}=0$ (writing x for χ^2)

or, $\left(\frac{n}{2}-1\right)-\frac{x}{2}=0$

or, $x=n-2$.

It can be shown that $f''(n-2)<0$.

So $f(\chi^2)$ has a unique maximum at $\chi^2=n-2$, when $n>2$ and so the $\chi^2(n)$ distribution has the unique mode M given by $M=n-2$ if $n>2$.

We know that the characteristic function $\phi(t)$ of a $\gamma(l)$ variate is given by

$$\phi(t)=(1-it)^{-l}. \tag{9.2.3}$$

Now if X is $\chi^2(n)$ variate, then $\frac{X}{2}$ is a $\gamma\left(\frac{n}{2}\right)$ variate and so the characteristic function of $\frac{X}{2}$ is given by

$$E(e^{it\frac{X}{2}})=(1-it)^{-\frac{n}{2}}, \quad \text{for all real } t.$$

Then $E(e^{itX})=(1-2it)^{-\frac{n}{2}}$, for all real t. (9.2.4)

(9.2.4) shows that the characteristic function $\phi_X(t)$ of X, *i.e.*, the characteristic function $\phi_X(t)$ of a $\chi^2(n)$ variate is given by

$$\phi_X(t)=(1-2it)^{-\frac{n}{2}}. \tag{9.2.5}$$

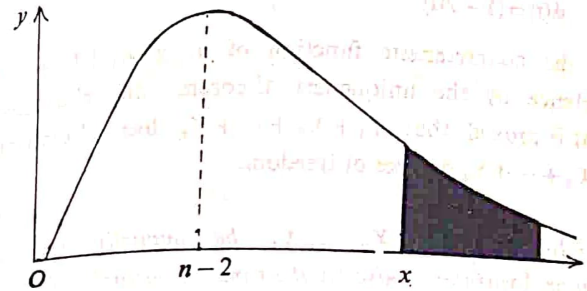


Fig. 9.2.1 Graph of the density curve of a $\chi^2(n)$ variate.

If χ^2 has $\chi^2(n)$ distribution, then $P(\chi^2>x)$ gives the area of the region shaded in Fig. 9.2.1. Table II (in the appendix) gives the values of x for different values of $P(\chi^2>x)$ and the number of degrees of freedom n. Value of x corresponding to $P(\chi^2>x)=\varepsilon(0<\varepsilon<1)$ and n degrees of freedom, will be denoted as $\chi^2_{\varepsilon,n}$. From table II (Appendix) we find that

$\chi^2_{0\cdot005,\,20}=39\cdot9968$,
$\chi^2_{0\cdot01,\,20}=37\cdot5662$,
$\chi^2_{0\cdot025,\,30}=46\cdot9792$ etc.

THEOREM 9.2.3. *If $X_1, X_2, \ldots, X_n$ are mutually independent χ^2-variates with degrees of freedom $k_1, k_2, \ldots, k_n$ respectively, then $X_1+X_2+\cdots+X_n$ has χ^2-distribution with $k_1+k_2+\cdots+k_n$ degrees of freedom.*

Proof: The characteristic function $\phi_{X_r}(t)$ of X_r is given by

$$\phi_{X_r}(t)=(1-2it)^{-\frac{k_r}{2}}, \quad r=1, 2, \ldots, n.$$

Let $S_n=X_1+X_2+\cdots+X_n$. If $\phi(t)$ be the characteristic function of S_n, then

$$\phi(t)=E\{e^{it(X_1+X_2+\cdots+X_n)}\}$$
$$=E(e^{itX_1})E(e^{itX_2})\ldots E(e^{itX_n}),$$

since $X_1, X_2, \ldots, X_n$ are mutually independent.

or, $\phi(t)=(1-2it)^{-\frac{k_1}{2}}(1-2it)^{-\frac{k_2}{2}}\dots(1-2it)^{-\frac{k_n}{2}}$

So the characteristic function of S_n is given by

$$\phi(t)=(1-2it)^{-\frac{k_1+k_2+\dots+k_n}{2}},$$

which is the characteristic function of a $\chi^2(k_1+k_2+\dots+k_n)$ variate. Hence by the uniqueness theorem on characteristic function, it is proved that $X_1+X_2+\dots+X_n$ has χ^2-distribution with $k_1+k_2+\dots+k_n$ degrees of freedom.

THEOREM 9.2.4. *Let $X_1, X_2, \dots, X_n$ be mutually independent standard normal variates. Also let the random variables $Y_1, Y_2, \dots, Y_m$ $(m<n)$ be defined by $Y_r=a_{r1}X_1+a_{r2}X_2+\dots+a_{rn}X_n$, for $r=1, 2, \dots, m$, where the real constants a_{rj} $(r=1, 2, \dots, m;\ j=1, 2, \dots, n)$ satisfy the conditions*

$$\sum_{j=1}^{n} a_{rj}\,a_{sj}=0 \quad \text{if } r\neq s$$
$$=1 \quad \text{if } r=s, \qquad (9.2.6)$$

where $r, s \in \{1, 2, \dots, m\}$.

Then $(X_1^2+X_2^2+\dots+X_n^2)-(Y_1^2+Y_2^2+\dots+Y_m^2)$ has χ^2 distribution with $n-m$ degrees of freedom.

Proof: Let $\alpha_r=(a_{r1}, a_{r2}, \dots, a_{rn})$, for $r=1, 2, \dots, m$. Then by (9.2.6), $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ is an orthonormal subset of n-dimensional Euclidean space R^n. So $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ can be extended to an orthonormal basis of R^n (Article 1.7) and so an orthogonal matrix A of size $n\times n$ can be found where $\alpha_1, \alpha_2, \dots, \alpha_m$ form the first m rows of A. Let the remaining $n-m$ rows of A be given by the vectors $\alpha_{m+1}, \dots, \alpha_n$, where

$$\alpha_{m+i}=(a_{m+i\,1}, a_{m+i\,2}, \dots, a_{m+i\,n}) \quad \text{for } i=1, 2, \dots, n-m.$$

Then we have $AA^T=A^TA=I_n$, $\quad\dots\quad$ (9.2.7)

where I_n is the unit matrix of size $n\times n$.

Now (9.2.7) gives

$$\sum_{j=1}^{n} a_{rj}\,a_{sj}=0 \text{ if } r\neq s$$
$$=1 \text{ if } r=s$$

and
$$\sum_{k=1}^{n} a_{kr}\,a_{ks}=0 \text{ if } r\neq s$$
$$=1 \text{ if } r=s$$

where $r, s \in \{1, 2, \dots, n\}$. (9.2.8)

We now define $n-m$ random variables $Y_{m+1}, Y_{m+2}, \dots, Y_n$ by

$$Y_{m+i}=a_{m+i\,1}X_1+\dots+a_{m+i\,n}X_n \quad \text{for } i=1, 2, \dots, n-m.$$

Then we get n random variables $Y_1, Y_2, \dots, Y_n$ given by

$$\begin{aligned} Y_1&=a_{11}X_1+a_{12}X_2+\dots+a_{1n}X_n\\ Y_2&=a_{21}X_1+a_{22}X_2+\dots+a_{2n}X_n\\ &\dots \quad \dots \quad \dots \quad \dots\\ Y_n&=a_{n1}X_1+a_{n2}X_2+\dots+a_{nn}X_n \end{aligned} \qquad (9.2.9)$$

where the coefficients a_{ij} $(i=1, 2, \dots, n;\ j=1,2,\dots, n)$ satisfy (9.2.8).

In real variables the transformation corresponding to (9.2.9) is giuen by

$$y_i=a_{i1}x_1+a_{i2}x_2+\dots+a_{in}x_n \quad \text{for } i=1, 2, \dots, n.$$

Then $\dfrac{\partial(y_1, y_2, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)}$

$$=\begin{vmatrix} a_{11} & a_{12}\dots\dots & a_{1n}\\ a_{21} & a_{22}\dots\dots & a_{2n}\\ \dots & \dots & \dots\\ \dots & \dots & \dots\\ a_{n1} & a_{n2}\dots\dots & a \end{vmatrix} = \det A=\pm 1,$$

since A is an orthogonal matrix.

So $\dfrac{\partial(y_1, y_2, \dots, y_n)}{\partial(x_1, x_2, \dots, x_n)}$ is either >0 or <0, for all real values of $x_1, x_2, \dots, x_n$.

Now the joint probability density function of $X_1, X_2, \cdots, X_n$ is given by

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(x_1^2+x_2^2+\dots+x_n^2)}.$$

Again, $y_1^2+y_2^2+\cdots+y_n^2 = x_1^2+x_2^2+\cdots+x_n^2$, by (9.2.8).

So, $f(x_1, x_2, \cdots, x_n) = \dfrac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(y_1^2+y_2^2+\cdots+y_n^2)}$.

Then if $\psi(y_1, y_2, \cdots, y_n)$ be the joint probability density function of $Y_1, Y_2, \cdots, Y_n$ we have

$$\psi(y_1, y_2, \dots, y_n) = \frac{1}{\left|\frac{\partial(y_1, y_2, \cdots, y_n)}{\partial(x_1, x_2, \cdots, x_n)}\right|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(y_1^2+y_2^2+\cdots+y_n^2)}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(y_1^2+y_2^2+\cdots+y_n^2)}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{y_n^2}{2}} \tag{9.2.10}$$

(9.2.10) shows that $Y_1, Y_2, \dots, Y_n$ are mutually independent standard normal variates. So $Y_{m+1}, Y_{m+2}, \dots, Y_n$ are also mutually independent normal variates.

$\therefore$ by Theorem 9.2.3 $Y_{m+1}^2+Y_{m+2}^2+\cdots+Y_n^2$ has $\chi^2(n-m)$ distribution. Again by (9.2.8),

$$Y_1^2+Y_2^2+\cdots+Y_n^2 = X_1^2+X_2^2+\cdots+X_n^2.$$

So, $Y_{m+1}^2+Y_{m+2}^2+\cdots+Y_n^2$

$$=(X_1^2+X_2^2+\cdots+X_n^2)-(Y_1^2+Y_2^2+\cdots+Y_m^2).$$

Hence it is proved that

$$(X_1^2+X_2^2+\cdots+X_n^2)-(Y_1^2+Y_2^2+\cdots+Y_m^2)$$

has χ^2-distribution with $n-m$ degrees of freedom.

9.3. t-Distribution.

The *t-distribution*, also called Student's distribution, is a continuous distribution whose probability density function $f(t)$ is given by

$$f(t) = \frac{1}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < \infty, \tag{9.3.1}$$

where n is a positive integer called the number of degrees of freedom of the distribution.

Mean, Median, Mode, Variance of the t-Distribution.

Let X be the random variable having t-distribution with n degrees of freedom.

For $n=1$, $f(t) = \dfrac{1}{\pi(1+t^2)}$, $-\infty < t < \infty$, and in this case $\int_{-\infty}^{\infty} tf(t)\,dt$ becomes $\int_{-\infty}^{\infty} \dfrac{t\,dt}{\pi(1+t^2)}$ which is not convergent and so $E(X)$, *i.e.*, mean does not exist for the t-distribution with one degree of freedom.

For $n > 1$, $E(X)$ exists if $\displaystyle\int_{-\infty}^{\infty} \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ is absolutely convergent. Again, $\displaystyle\int_{-\infty}^{\infty} \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ is absolutely convergent if

$\displaystyle\int_{a}^{\infty} \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ is convergent and the last integral is convergent if

$\displaystyle\int_{a}^{\infty} \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ is convergent $(a > 0)$.

Now taking $\phi(t) = \dfrac{t}{t^{n+1}} = \dfrac{1}{t^n}$ and $f_1(t) = \dfrac{t}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$,

we find that $\lim\limits_{t\to\infty} \frac{f_1(t)}{\phi(t)} = \underset{t\to\infty}{Lt} \frac{t^{n+1}}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$

$$= \underset{t\to\infty}{Lt} \frac{1}{\left(\frac{1}{t^2}+\frac{1}{n}\right)^{\frac{n+1}{2}}}$$

$= n^{\frac{n+1}{2}}$ which is finite (> 0).

Also $\int_a^\infty \phi(t)\,dt = \int_a^\infty \frac{dt}{t^n}$ is convergent if $n > 1$.

So by limit form of comparison test, $\int_a^\infty f_1(t)\,dt$,

i.e., $\int_a^\infty \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ is convergent if $n > 1$.

Hence, $\int_{-\infty}^\infty \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ is absolutely convergent if $n > 1$.

So, $E(X)$ exists if $n > 1$ and for $n > 1$,

$$E(X) = \frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)} \int_{-\infty}^\infty \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \frac{1}{\sqrt{n}\,B\left(\frac{1}{2},\frac{n}{2}\right)} \underset{B_1\to\infty}{Lt} \int_{-B_1}^{B_1} \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} = 0,$$

since $\frac{t}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ being an odd function, $\int_{-B_1}^{B_1} \frac{t\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} = 0$ for all B_1.

So the mean of the t-distribution with n-degrees of freedom is 0 if $n > 1$.

We observe that

$\int_{-\infty}^0 f(t)\,dt = \frac{1}{2}$, $\int_{-\infty}^x f(t)\,dt < \frac{1}{2}$ if $x < 0$, $\int_{-\infty}^x f(t)\,dt > \frac{1}{2}$ if $x > 0$, where $f(t)$ is given by (9.3.1). So $\mu = 0$ is the unique solution of

$\int_{-\infty}^{\mu} f(t)\,dt = \frac{1}{2}$. Hence the median of the t-distribution with n degrees of freedom is 0 for every n.

Further we observe that for any n, $f'(t) = 0$ implies $t = 0$ and $f''(0) < 0$. So 0 is the unique mode of the t-distribution for every n.

Thus the mean (when it exists), median, mode of the t-distribution are all equal to 0, as expected for a symmetrical distribution. [Here the distribution is symmetrical since the probability density function $f(t)$ satisfies $f(0+t) = f(0-t)$, for all t.]

Now let us find the variance of t-distribution with n-degrees of freedom. We have seen that the mean does not exist for $n = 1$. So the variance of the t-distribution also does not exist for $n = 1$. Also we have shown that the mean of the t-distribution is 0 if $n > 1$. Then var $X = E(X^2)$ provided $E(X^2)$ exists where X has t distribution with n degrees of freedom.

Now $E(X^2)$ will exist if

$$\int_{-\infty}^\infty \frac{t^2\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \text{ is absolutely convergent.}$$

Again, $\int_{-\infty}^\infty \frac{t^2\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$ will be absolutely convergent if

$$\int_a^\infty \frac{t^2\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \text{ is convergent } (a > 0).$$

Now by the limit form of comparison test we find that

$$\int_a^\infty \frac{t^2\,dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}} \text{ is convergent iff } \int_a^\infty \frac{dt}{t^{n-1}}$$

is convergent and we know that $\int_a^\infty \frac{dt}{t^{n-1}}$ $(a > 0)$ is convergent iff $n-1 > 1$, i.e., iff $n > 2$.

So variance of the t-distribution exists iff $n > 2$ and for such n,

$$\text{var } X = E(X^2) = \frac{1}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{-\infty}^{\infty} \frac{t^2\, dt}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$$= \frac{n\sqrt{n}}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\tan^2\theta \sec^2\theta\, d\theta}{\sec^{n+1}\theta}, \quad \text{where } t = \sqrt{n}\tan\theta$$

$$= \frac{n}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2\theta \cos^{n-3}\theta\, d\theta$$

$$= \frac{2n}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \int_{0}^{\frac{\pi}{2}} \sin^2\theta \cos^{n-3}\theta\, dp$$

$$= \frac{2n}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{2} \frac{\Gamma(\frac{3}{2})\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}$$

$$= \frac{n\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{\frac{1}{2}\Gamma(\frac{1}{2})\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}$$

$$= \frac{\frac{n}{2}\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \frac{\frac{n}{2}\Gamma\left(\frac{n-2}{2}\right)}{\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} = \frac{n}{n-2}.$$

Thus it is proved that variance of the t-distribution (with n degrees of freedom exists iff $n > 2$) and is equal to $\frac{n}{n-2}$.

It is symmetrical about $t = 0$. It can be shown that $\gamma_2 > 0$ and so the graph of $f(t)$ has a more sharp peak than the normal distribution with the same standard deviation.

If X has t distribution with n degrees of freedom, then the value $P(X > t) = \varepsilon$, (say) $(0 < \varepsilon < 1)$ gives the area of the region shaded in Fig. 9.3.1. Value of t satisfying $P(X > t) = \varepsilon$, will be denoted as $t_{\varepsilon, n}$. Table III (appendix) gives the values of t for different values of

$P(X > t)$ and the number of degrees of freedom n. From this table we find that $t_{0.4,\,15} = 0.258$, $t_{0.05,\,20} = 1.725$ etc.

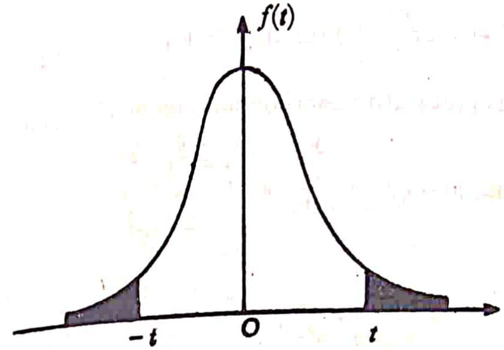


Fig. 9.3.1. Graph of the probability density function of t-distribution.

THEOREM 9.3.1. *If X be a standard normal variate and Y has χ^2 distribution with n degrees of freedom and if X, Y are independent, then $\dfrac{X}{\sqrt{\frac{Y}{n}}}$ has t-distribution with n degrees of freedom.*

Proof: Since X, Y are respectively $N(0, 1)$ and $\chi^2(n)$ variates and X, Y are independent, the joint probability density function $f_{X,Y}(x, y)$ is given by

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n}{2}\right)}, \; -\infty < x < \infty,\; 0 < y < \infty.$$

Let $U = Y$, $V = \dfrac{X}{\sqrt{\frac{Y}{n}}}$.

In real variables we have $u = y$, $v = \dfrac{x}{\sqrt{\frac{y}{n}}}$,

from which we get $x = v\sqrt{\frac{u}{n}}$, $y = u$.

So $\dfrac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \dfrac{v}{\sqrt{n}}\dfrac{1}{2\sqrt{u}} & \sqrt{\dfrac{u}{n}} \\ 1 & 0 \end{vmatrix}$

$$= -\sqrt{\frac{u}{n}} < 0 \text{ for all } y > 0 \text{ and for all } x \in (-\infty, \infty).$$

So the joint probability density function of U, V is given by

$$f_{U,V}(u, v) = \sqrt{\frac{u}{n}} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, \frac{e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{1}{2\sqrt{2\pi}} \cdot \frac{\sqrt{u}}{\sqrt{n}}\, e^{-\frac{v^2 u}{2n}} \cdot \frac{e^{-\frac{u}{2}}\left(\frac{u}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{e^{-\left(\frac{u}{2}+\frac{uv^2}{2n}\right)} u^{\frac{n}{2}-\frac{1}{2}}}{\sqrt{2\pi} \cdot \sqrt{n}\,\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}$$

$$= \frac{u^{\frac{n-1}{2}} e^{-\frac{u}{2}\left(1+\frac{v^2}{n}\right)}}{\sqrt{2n\pi}\, 2^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)}.$$

So we get $f_{U,V}(u, v) = \dfrac{e^{-\frac{u}{2}\left(1+\frac{v^2}{n}\right)} \cdot u^{\frac{n-1}{2}}}{\sqrt{2n\pi}\, 2^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)}$,

for $0 < u < \infty$, $-\infty < v < \infty$.

Hence the probability density function $f_V(v)$ of V is given by

$$f_V(v) = \int_0^\infty \frac{e^{-\frac{u}{2}\left(1+\frac{v^2}{n}\right)} \cdot u^{\frac{n-1}{2}}}{\sqrt{2n\pi}\, 2^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)}\, du, \quad (-\infty < v < \infty)$$

$$= \frac{1}{\sqrt{2n\pi}\, 2^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-z}\left(\frac{2z}{1+\frac{v^2}{n}}\right)^{\frac{n-1}{2}} \frac{2\, dz}{1+\frac{v^2}{n}},$$

where $z = \frac{u}{2}\left(1+\frac{v^2}{n}\right)$

or, $f_V(v) = \dfrac{2^{\frac{n+1}{2}}}{\sqrt{2n\pi} \cdot 2^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)\left(1+\frac{v^2}{n}\right)^{\frac{n+1}{2}}} \displaystyle\int_0^\infty e^{-z} z^{\frac{n+1}{2}-1}\, dz$

$$= \frac{1}{\sqrt{n}\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{v^2}{n}\right)^{\frac{n+1}{2}}}\, \Gamma\left(\frac{n+1}{2}\right).$$

Now $B\left(\frac{1}{2}, \frac{n}{2}\right) = \dfrac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} = \dfrac{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}$.

$\therefore \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \dfrac{\sqrt{\pi}}{B\left(\frac{1}{2}, \frac{n}{2}\right)}$.

Hence, we get $f_V(v) = \dfrac{1}{\sqrt{n}\, B\left(\frac{1}{2}, \frac{n}{2}\right)} \dfrac{1}{\left(1+\frac{v^2}{n}\right)^{\frac{n+1}{2}}}$ for $-\infty < v < \infty$,

and this shows that V, *i.e.*, $\dfrac{X}{\sqrt{\frac{Y}{n}}}$ has t-distribution with n degrees of freedom.

9.4. F-Distribution.

A continuous random variable X is said to have *F-distribution* with parameters n_1, n_2 (positive integers) if the corresponding probability density function $f(F)$ (here the corresponding real variable is usually denoted as F instead x) is given by

$$f(F) = \frac{n_1^{\frac{n_1}{2}}\, n_2^{\frac{n_2}{2}}\, F^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(n_1 F + n_2\right)^{\frac{n_1+n_2}{2}}} \quad \text{if } F > 0$$

$$= 0 \quad \text{elsewhere.}$$

If X has F-distribution with parameters n_1, n_2, we say that X is an $F(n_1, n_2)$ variate.

We observe that if $f(F)$ and $f_1(F)$ are respectively the probability density functions of $F(n_1, n_2)$ and $F(n_2, n_1)$ variates, then $f(F) \neq f_1(F)$ (unless $n_1 = n_2$) and so the F-distribution is not symmetrical with respect to n_1, n_2.

THEOREM 9.4.1. *If X_1, X_2 are independent χ^2 variates with degrees of freedom n_1, n_2 respectively, then $\dfrac{X_1/n_1}{X_2/n_2}$ is an $F(n_1, n_2)$ variate.*

Proof: Let $Y = \dfrac{X_1/n_1}{X_2/n_2}$

Then we have $\dfrac{n_1}{n_2} Y = \dfrac{X_1}{X_2} = \dfrac{\frac{1}{2}X_1}{\frac{1}{2}X_2}$.

Now by Theorem 9.2.1, $\frac{1}{2}X_1, \frac{1}{2}X_2$ are respectively $\gamma\left(\frac{n_1}{2}\right), \gamma\left(\frac{n_2}{2}\right)$ variates. Also $\frac{1}{2}X_1, \frac{1}{2}X_2$ are independent.

Hence, (by illustrative Ex. 27, Chapter VI), $\dfrac{\frac{1}{2}X_1}{\frac{1}{2}X_2}$ is a $\beta_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ variate.

Then making the transformation $X = \dfrac{n_1}{n_2} Y$, we find that X is a $\beta_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ variate.

The above transformation can also be expressed as $Y = \dfrac{n_2}{n_1} X$.

In real variables we get $y = \dfrac{n_2}{n_1} x$.

Then $\dfrac{dy}{dx} = \dfrac{n_2}{n_1} > 0$ for all x. Then if $f_Y(y)$ be the probability density function of Y, then

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x), \text{ where}$$

$$f_X(x) = \frac{x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(1+x)^{\frac{n_1+n_2}{2}}} \text{ if } 0 < x < \infty$$
$$= 0, \qquad \text{elsewhere.}$$

$\left[\because\ X \text{ is a } \beta_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \text{ variate.}\right]$

So we get $$f_Y(y) = \frac{n_1}{n_2} \frac{x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(1+x)^{\frac{n_1+n_2}{2}}} \text{ if } 0 < x < \infty$$

$$= \frac{n_1}{n_2} \frac{\left(\frac{n_1 y}{n_2}\right)^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1 y}{n_2}\right)^{\frac{n_1+n_2}{2}}} \text{ if } 0 < y < \infty.$$

$$\therefore\ f_Y(y) = \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} y^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(n_1 y + n_2)^{\frac{n_1+n_2}{2}}} \text{ if } 0 < y < \infty.$$

Hence the probability density function of Y is given by

$$f_Y(y) = \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} y^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(n_1 y + n_2)^{\frac{n_1+n_2}{2}}} \text{ if } y > 0$$
$$= 0, \qquad \text{elsewhere.} \tag{9.4.1}$$

(9.4.1) shows that Y, *i.e.*, $\dfrac{X_1/n_1}{X_2/n_2}$ is an $F(n_1, n_2)$ variate.

THEOREM 9.4.2. *If X is an $F(n_1, n_2)$ variate, then $\dfrac{1}{X}$ is an $F(n_2, n_1)$ variate.*

Proof: Let $Y = \dfrac{1}{X}$.

In real variables we have $y = \dfrac{1}{x}$.

Then $\dfrac{dy}{dx} = -\dfrac{1}{x^2} < 0$ for all x.

So if $f_X(x)$ and $f_Y(y)$ be the probability density functions of X, Y respectively, we have $f_Y(y) = |x^2|\ f_X(x) = x^2 f_X(x)$.

$$\text{Now } f_X(x) = \frac{n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(n_1 x + n_2)^{\frac{n_1+n_2}{2}}} \text{ if } x > 0$$
$$= 0, \qquad \text{elsewhere.}$$

$$\text{Hence,}\quad f_Y(y) = \frac{1}{y^2}\,\frac{n_1^{\frac{n_1}{2}}\, n_2^{\frac{n_2}{2}} \left(\frac{1}{y}\right)^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(n_2 + \frac{n_1}{y}\right)^{\frac{n_1+n_2}{2}}} \quad \text{if } y > 0$$

$$= \frac{n_2^{\frac{n_2}{2}}\, n_1^{\frac{n_1}{2}}\, y^{\frac{n_2}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(n_2 y + n_1)^{\frac{n_1+n_2}{2}}} \quad \text{if } y > 0$$

Thus it is proved that

$$f_Y(y) = \frac{n_2^{\frac{n_2}{2}}\, n_1^{\frac{n_1}{2}}\, y^{\frac{n_2}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)(n_2 y + n_1)^{\frac{n_1+n_2}{2}}} \quad \text{if } y > 0$$

$$= 0 \quad \text{elsewhere.} \qquad (9.4.2)$$

(9.4.2) shows that $Y = \frac{1}{X}$ is an $F(n_2, n_1)$ variate.

We now state below (without proof) some important properties of the F-distribution :

(*i*) The mean of an $F(n_1, n_2)$ variate exists iff $n_2 > 2$ and its value is $\frac{n_2}{n_2 - 2}$.

(*ii*) The F-distribution is highly positively skewed.

(*iii*) $F(1, n_2)$ variate can be identified with a random variable X^2, where X has t-distribution with n_2 degrees of freedom.

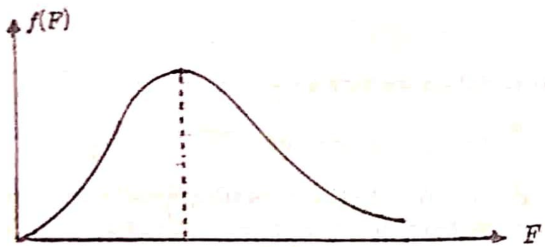


Fig. 9.4.1 Density Curve of F-distribution.

If $P(X > a) = \varepsilon$, $(0 < \varepsilon < 1)$ where X is an $F(n_1, n_2)$ variate, we denote the value of a as $F_{\varepsilon;\, n_1, n_2}$

Then using the above notation we can write

$$P(X > F_{\varepsilon;\, n_1, n_2}) = \varepsilon$$

$$\text{or,}\quad 1 - P(X \leqslant F_{\varepsilon;\, n_1, n_2}) = \varepsilon$$

$$\text{or,}\quad P(X < F_{\varepsilon;\, n_1, n_2}) = 1 - \varepsilon \qquad (9.4.3)$$

$$\text{or,}\quad P\left(\frac{1}{X} > \frac{1}{F_{\varepsilon;\, n_1, n_2}}\right) = 1 - \varepsilon. \qquad (9.4.4)$$

But by the Theorem 9.4.2, $\frac{1}{X}$ is an $F(n_2, n_1)$ variate. Then (9.4.4) shows that $\frac{1}{F_{\varepsilon;\, n_1, n_2}}$ can be denoted as $F_{1-\varepsilon;\, n_2, n_1}$. So we get the relation

$$F_{\varepsilon;\, n_1, n_2} = \frac{1}{F_{1-\varepsilon;\, n_2, n_1}}. \qquad (9.4.5)$$

Table IV (appendix) gives the values of $F_{\varepsilon;\, n_1, n_2}$ for different values of n_1, n_2 and for $\varepsilon = 0.01, 0.05$. For example, from the table IV we find that

$$F_{0.05;\, 8, 10} = 3.07,$$

$$F_{0.01;\, 8, 10} = 5.06.$$

9.5. Illustrative Examples.

Ex. 1. *The mean I. Q of a group of children is 90 with a standard deviation of 20. Assuming that I. Q. is normally distributed, find the percentage of children with I. Q. over 100. [Given $\Phi(0.5) = 0.6915$ where $\Phi(x)$ is the distribution function of standard normal variate.]*

Let X be the random variable denoting the I. Q. of the given group of children. Then X is a normal variate with mean $m = 90$ and standard deviation $\sigma = 20$.

So $U = \frac{X - 90}{20}$ is a standard normal variate.

$$\text{Now}\quad P(X > 100) = P\left(\frac{X-90}{20} > \tfrac{1}{2}\right) = P(U > 0.5)$$

$$= 1 - P(U \leqslant 0.5) = 1 - \Phi(0.5) = 0.3085.$$

Hence the percentage of students having *I. Q* over 100 is

$$0.3085 \times 100 = 30.85.$$

Ex. 2. *5000 candidates appeared at an examination, in which the minimum for a pass is 40 and the minimum for a distinction is 50. It is known that the average marks obtained by the candidates is 43 and standard deviation is 7. Find how many candidates expect to get (i) simply pass, (ii) distinction. Assume that the distribution of marks is normal.* [*Given* $\Phi(0.43)=0.6664$, $\Phi(1)=0.8413$.]

Let X be the random variable denoting the marks of the given candidates. Then X is a normal variate with mean $m=43$ and standard deviation $\sigma=7$.

So $U=\dfrac{X-43}{7}$ is a standard normal variate.

Now the probability that a candidate gets simply 'pass' is

$$\begin{aligned}
&P(40 \leqslant X < 50)\\
&=P(-3 \leqslant X-43 < 7)\\
&=P(-\tfrac{3}{7} \leqslant U < 1)\\
&=P(-0.43 \leqslant U < 1)\\
&=\Phi(1)-\Phi(-0.43)\\
&=\Phi(1)-\{1-\Phi(0.43)\}\\
&=\Phi(1)+\Phi(0.43)-1\\
&=0.8413+0.6664-1\\
&=1.5077-1\\
&=0.5077.
\end{aligned}$$

So the number of candidates (out of 5000 candidates) expecting to get simply 'pass' is approximately $5000\times 0.5077=2538.5\simeq 2538$.

Again the probability that a candidate gets distinction is $P(X \geqslant 50)$

$$\begin{aligned}
&=P\left(\frac{X-43}{7}\geqslant 1\right)=P(U\geqslant 1)\\
&=1-P(U<1)=1-P(U\leqslant 1)\\
&=1-\Phi(1)=1-0.8413=0.1587.
\end{aligned}$$

So the number of candidates expecting to get distinction is

$$5000\times 0.1587=793.5\simeq 794.$$

Ex. 3. *In an examination, marks obtained by the students in Mathematics, Physics and Chemistry are independently and normally distributed with means 50, 52, 48 and standard deviations 15, 12, 16 respectively. Find the probability that the total marks of a student are (i) 180 or more, (ii) 90 or less.*

[*Given* $\Phi(1.2)=0.8849$, $\Phi(2.4)=0.9918$.]

Let X, Y, Z be the random variables denoting respectively the marks in Mathematics, Physics and Chemistry. Then X, Y, Z are mutually independent normal variates with parameters (50, 15), (52, 12) and (48, 16) respectively. Hence, by the reproductive property of normal distribution $X+Y+Z$ is a normal variate with mean $50+52+48=150$ and standard deviation $\sqrt{(15)^2+(12)^2+(16)^2}=25$.

So $U=\dfrac{X+Y+Z-150}{25}$ is a standard normal variate.

Now $X+Y+Z$ is the random variable denoting the total marks of a student. We are to find the values of $P(X+Y+Z\geqslant 180)$ and $P(X+Y+Z\leqslant 90)$.

Now $P(X+Y+Z\geqslant 180)$

$$\begin{aligned}
&=P\left(\frac{X+Y+Z-150}{25}\geqslant\frac{30}{25}\right)=P(U\geqslant 1.2)\\
&=1-P(U<1.2)=1-P(U\leqslant 1.2)\\
&=1-\Phi(1.2)=1-0.8849=0.1151.
\end{aligned}$$

Again, $P(X+Y+Z\leqslant 90)=P\left(\dfrac{X+Y+Z-150}{25}\leqslant\dfrac{-60}{25}\right)$

$$\begin{aligned}
&=P(U\leqslant -2.4)=\Phi(-2.4)\\
&=1-\Phi(2.4)=1-0.9918=0.0082.
\end{aligned}$$

Ex. 4. *If X has uniform distribution in $(0, a)$, then find the distribution of* $-2\log_e\dfrac{X}{a}$.

Let $Y=-2\log_e\dfrac{X}{a}$.

In real variables we have

$$y=-2\log_e\frac{x}{a},\quad x>0$$

Then $\dfrac{dy}{dx}=-\dfrac{2}{x}<0$ for all $x>0$.

So the probability density function $f_Y(y)$ of Y is given by

$f_Y(y)=\left|-\dfrac{x}{2}\right| f_X(x)$, where

$$\begin{aligned}
f_X(x)&=\frac{1}{a} &&\text{if } 0<x<a\\
&=0, &&\text{elsewhere.}
\end{aligned}$$

$$\therefore \quad f_Y'(y) = \frac{1}{2a} a e^{-\frac{y}{2}} \quad (\because \; x = a e^{-\frac{y}{2}})$$

$$= \tfrac{1}{2} e^{-\frac{y}{2}} \text{ if } y > 0.$$

Hence the probability density function of Y is given by

$$f_Y'(y) = \frac{e^{-\frac{y}{2}} \left(\frac{y}{2}\right)^{\frac{2}{2}-1}}{\Gamma\left(\frac{2}{2}\right)} \text{ if } y > 0.$$

$$= 0, \qquad \text{elsewhere,}$$

and this shows that Y, *i.e.*, $-2 \log_e \frac{X}{a}$ has χ^2 distribution with 2 degrees of freedom.

Ex 5. *If X, Y are independent χ^2 variates with degrees of freedom m, n respectively, then find the joint distribution of $X+Y$ and $\cot^{-1} \sqrt{\frac{X}{Y}}$ and hence, prove that $X+Y$ is a $\chi^2(m+n)$ variate.*

Let $\sqrt{X} = \sqrt{U} \cos V$,

$\sqrt{Y} = \sqrt{U} \sin V$.

Then $X+Y = U$, $\cot^2 V = \frac{X}{Y}$.

In real variables, $x = u \cos^2 v$, $y = u \sin^2 v$.

$$\text{Then} \quad \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \cos^2 v & -2u \sin v \cos v \\ \sin^2 v & 2u \sin v \cos v \end{vmatrix}$$

$$= 2u \sin v \cos v = u \sin 2v > 0$$

for all $u > 0$ and $0 < v < \frac{\pi}{2}$.

Then $f_{U,V}(u, v) = u \sin 2v \, f_{X,Y}(x, y)$

$$= u \sin 2v \, f_X(x) f_Y(y)$$

$$= \sin 2v \frac{e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{m}{2}-1} e^{-\frac{y}{2}} \left(\frac{y}{2}\right)^{\frac{n}{2}-1}}{2\Gamma\left(\frac{m}{2}\right) \cdot 2\Gamma\left(\frac{n}{2}\right)} \text{ if } x > 0, y > 0$$

$$= \frac{u \sin 2v \, e^{-\frac{u}{2}} (u \cos^2 v)^{\frac{m}{2}-1} (u \quad ^2 v)^{\frac{n}{2}-}}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)}$$

if $u > 0$, $0 < v < \frac{\pi}{2}$.

$$\text{So} \quad f_{U,V}(u, v) = \frac{2u^{\frac{m+n}{2}-1} e^{-\frac{u}{2}}}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m+n}{2}\right) B\left(\frac{m}{2}, \frac{n}{2}\right)} \cos^{m-1} v \sin^{n-1} v$$

$$= \frac{e^{-\frac{u}{2}} \left(\frac{u}{2}\right)^{\frac{m+n}{2}-1}}{2\Gamma\left(\frac{m+n}{2}\right)} \cdot \frac{2 \sin^{n-1} v \cos^{m-1} v}{B\left(\frac{m}{2}, \frac{n}{2}\right)}$$

if $u > 0$, $0 < v < \frac{\pi}{2}$

which gives the joint distribution of U and V, *i.e.*, of $X+Y$ and $\cot^{-1} \sqrt{\frac{X}{Y}}$.

Then the marginal probability density function of U is given by

$$f_U(u) = \frac{e^{-\frac{u}{2}} \left(\frac{u}{2}\right)^{\frac{m+n}{2}-1}}{2\Gamma\left(\frac{m+n}{2}\right)} \int_0^{\frac{\pi}{2}} \frac{2 \sin^{n-1} v \cos^{m-1} v \, dv}{B\left(\frac{m}{n}, \frac{n}{2}\right)}$$

$$= \frac{e^{-\frac{u}{2}} \left(\frac{u}{2}\right)^{\frac{m+n}{2}-1}}{2\Gamma\left(\frac{m+n}{2}\right)} \text{ if } u > 0,$$

which shows that U, *i.e.*, $X+Y$ is a $\chi^2(m+n)$ variate.

Ex. 6 *Let X_μ be a Poisson variate with parameter μ and Y_k be a $\chi^2(k)$ variate. Prove that*

$$P(X_\mu \leqslant k-1) = P(Y_{2k} > 2\mu),$$

for all positive integer k.

We have $P(X_\mu \leqslant k-1) = \sum_{r=0}^{k-1} \frac{e^{-\mu} \mu^r}{r!}$.

Now Y_{2k} is a $\chi^2(2k)$ variate.

Then $P(Y_{2k} > 2\mu)$

$$= \int_{2\mu}^{\infty} \frac{e^{-\frac{u}{2}} \left(\frac{u}{2}\right)^{\frac{2k}{2}-1}}{2\Gamma\left(\frac{2k}{2}\right)} du = \int_{2\mu}^{\infty} \frac{e^{-\frac{u}{2}} \left(\frac{u}{2}\right)^{k-1}}{2\Gamma(k)} du$$

$$= \frac{1}{2^k (k-1)!} \int_{2\mu}^{\infty} e^{-\frac{u}{2}} u^{k-1} du$$

or, $P(Y_{2k} > 2\mu) = \frac{1}{2^k(k-1)!}\int_{\mu}^{\infty} e^{-v}(2v)^{k-1} \cdot 2dv$. where $\frac{u}{2} = v$

$$= \frac{1}{(k-1)!}\int_{\mu}^{\infty} e^{-y}y^{k-1}dy$$

$= P(X_\mu \leqslant k-1)$, by (See Ex. 47 of § 5·14)

Hence it is proved that

$$P(X_\mu \leqslant k-1) = P(Y_{2k} > 2\mu),$$

for all positive integer k.

Ex. 7. *Let X, Y, Z, U be mutually independent standard normal variates Find the value of $P(X^2 - 99Y^2 \leqslant 99Z^2 - U^2)$.*

[Given $F_{0\cdot 1\,;\,2,2} = 99$.]

$P(X^2 - 99Y^2 \leqslant 99Z^2 - U^2)$

$= P\{X^2 + U^2 \leqslant 99(Y^2 + Z^2)\}$

Now $X^2 + U^2$ and $Y^2 + Z^2$ are both $\chi^2(2)$ variates and they are independent. So by the theorem 9.4.1, $\dfrac{\frac{X^2+U^2}{2}}{\frac{Y^2+Z^2}{2}}$ is an $F(2, 2)$ variate,

i.e., $\dfrac{X^2+U^2}{Y^2+Z^2}$ is an $F(2, 2)$ variate.

Then the required probability is equal to

$$P\left(\frac{X^2+U^2}{Y^2+Z^2} \leqslant 99\right) = P(F \leqslant 99),$$

where $F = \dfrac{X^2+U^2}{Y^2+Z^2}$ is an $F(2, 2)$ variate.

Now it is given that $F_{0\cdot 01\,;\,2,\,2} = 99$.

So $P(F \leqslant 99) = 1 - P(F > 99) = 1 - 0{\cdot}01 = 0{\cdot}99$,

where F is an F (2, 2) variate.

Hence, $P\left(\dfrac{X^2+U^2}{Y^2+Z^2} \leqslant 99\right) = 0{\cdot}99$.

So the required probability is 0·99.

Ex. 8. *Let X_1, X_2, X_3 be three mutually independent standard normal variates. Find the value of $P\{2X_1^2 \leqslant (6{\cdot}965)^2(X_2^2 + X_3^2)\}$.*

[Given $t_{0\cdot 01\,;\,2} = 6{\cdot}965$.]

Here $X_2^2 + X_3^2$ is a $\chi^2(2)$ variate and X_1 is a standard normal variate and they are independent. Then by the theorem 9.3.1,

$\dfrac{X_1}{\sqrt{\dfrac{X_2^2+X_3^2}{2}}}$ has t-disiribution with 2 degrees of freedom.

Let $T = \dfrac{X_1}{\sqrt{\dfrac{X_2^2+X_3^2}{2}}}$.

Now the required probability is

$P\{2X_1^2 \leqslant (6{\cdot}965)^2(X_2^2 + X_3^2)\}$

$= P\left\{\dfrac{X_1^2}{\frac{X_2^2+X_3^2}{2}} \leqslant (6{\cdot}965)^2\right\}$

$= P\{T^2 \leqslant (6{\cdot}965)^2\}$

$= P(|T| \leqslant 6{\cdot}965)$, where T has t-distribution with 2 degrees of freedom.

Now $P(|T| \leqslant 6{\cdot}965)$

$= 1 - P(|T| > 6{\cdot}965)$

$= 1 - 2\,P(T > 6{\cdot}965)$, due to symmetry of the t-distribution

$= 1 - 2 \times 0{\cdot}01$

$= 0{\cdot}98$.

Hence,the required probability is 0·98.

Ex. 9. *Assume that the velocity components V_x, V_y, V_z of any molecule of a gas are mutually independent random variables, each being normal $\left(0, \sqrt{\frac{kT}{m}}\right)$ where k is Boltzmann's constant, T the absolute temperature of the gas and m the mass of a molecule. Find the probability density function of the velocity*

$$V = \sqrt{V_x^2 + V_y^2 + V_z^2}.$$

Here V_x, V_y, V_z are mutually independent normal $\left(0, \sqrt{\frac{kT}{m}}\right)$ variates. So

$$\frac{V_x}{\sqrt{\frac{kT}{m}}}, \frac{V_y}{\sqrt{\frac{kT}{m}}}, \frac{V_z}{\sqrt{\frac{kT}{m}}}$$ are mutually independent standard normal variates. Hence,

$$\frac{V_x^2+V_y^2+V_z^2}{\frac{kT}{m}}$$ is a $\chi^2(3)$ variate.

Here $V^2=V_x^2+V_y^2+V_z^2$. So $\dfrac{V^2}{\frac{kT}{m}}$ is a $\chi^2(3)$ variate.

Let $X=\dfrac{V^2}{\frac{kT}{m}}$. Then $V=\sqrt{\dfrac{kT}{m}}\sqrt{X}$ where X is a $\chi^2(3)$ variate.

Transformation in real variables corresponding to

$$V=\sqrt{\frac{kT}{m}}\sqrt{X} \text{ is } v=\sqrt{\frac{kT}{m}}\sqrt{x}.$$

Then $\dfrac{dv}{dx}=\dfrac{1}{2\sqrt{x}}\sqrt{\dfrac{kT}{m}}>0$ for all x. So if $f(v)$ and $\varphi(x)$ be the probability density functions of V and X respectively, then

$$f(v)=2\sqrt{x}\sqrt{\frac{m}{kT}}\,\varphi(x).$$

Now, $$\varphi(x)=\frac{e^{-\frac{x}{2}}\left(\frac{x}{2}\right)^{\frac{3}{2}-1}}{2\Gamma(\frac{3}{2})} \quad \text{if } x>0$$
$$=0 \quad \text{elsewhere.}$$

So $$f(v)=\frac{\sqrt{\frac{m}{kT}}\,v\,e^{-\frac{1}{2}v^2\frac{m}{kT}}\left(\frac{v}{\sqrt{2}}\sqrt{\frac{m}{kT}}\right)\sqrt{\frac{m}{kT}}}{\Gamma(\frac{3}{2})} \quad \text{if } v>0$$

$$=\frac{v^2e^{-\beta v^2}}{\sqrt{2}\,\frac{1}{2}\Gamma(\frac{1}{2})}\left(\frac{m}{kT}\right)^{\frac{3}{2}} \quad \text{if } v>0$$

$$=\frac{v^2e^{-\beta v^2}}{\frac{1}{\sqrt{2}}\sqrt{\pi}}\sqrt{\frac{\pi}{2}}\,\alpha \quad \text{if } v>0$$

where $\alpha=\sqrt{\dfrac{2}{\pi}}\left(\dfrac{m}{kT}\right)^{\frac{3}{2}}$, $\beta=\dfrac{m}{2kT}$

$$\therefore\ f(v)=\alpha\cdot v^2e^{-\beta v^2} \quad \text{if } v>0$$
$$=0 \quad \text{elsewhere,}$$

which gives the distribution of V.

Ex. 10. *If X, Y, Z are mutually independent standard normal variates, then find the values of*

(i) $P(X-2Y+2Z \leq 3)$,

(ii) $P(X+2Y+2Z \geq 5{\cdot}88)$,

(iii) $P(X^2+Z^2 \geq 0{\cdot}584-Y^2)$.

[*Given* $\Phi(1)=0{\cdot}8413$, $\Phi(1{\cdot}96)=0{\cdot}9750$, $\chi^2_{0{\cdot}90,\,3}=0{\cdot}584$.]

(i) Here $X-2Y+2Z$ has normal distribution with mean 0 and standard deviation $\sqrt{1+4+4}=3$.

So $U=\dfrac{X-2Y+2Z}{3}$ is a standard normal variate.

$$\therefore\ P(X-2Y+2Z \leq 3)$$
$$=P(U\leq 1)$$
$$=\Phi(1)={\cdot}8413.$$

(ii) Here $X+2Y+2Z$ has normal distribution with mean 0 and standard deviation $\sqrt{1+4+4}=3$.

So $U=\dfrac{X+2Y+2Z}{3}$ is a standard normal variate.

Now $P(X+2Y+2Z \geq 5{\cdot}88)$.

$$=P\left(\frac{P+2Y+2Z}{3}\geq\frac{5{\cdot}88}{3}\right)$$
$$=P(U\geq 1{\cdot}96)$$
$$=1-P(U<1{\cdot}96)$$
$$=1-P(U\leq 1{\cdot}96)$$
$$=1-\Phi(1{\cdot}96)$$
$$=1-0{\cdot}9750$$
$$=0{\cdot}0250.$$

(*iii*) Here $X^2+Y^2+Z^2$ is a χ^2 (3) variate.

So $P(X^2+Z^2 > 0.584-Y^2)$

$= P(X^2+Y^2+Z^2 > 0.584)$

$= P(\chi^2 > 0.584)$,

where $\chi^2 = X^2+Y^2+Z^2$ has χ^2 (3) distribution.

Now it is given that $\chi^2_{0.090,\,3} = 0.584$.

$\therefore$ $P(X^2 \geqslant 0.584) = P(\chi^2 > 0.584) = 0.90$.

Hence the required probability is 0·90.

Examples IX

1. If $f(x)$ is the density function of a normal $(0, \sigma)$ variate, then show that $\int_{-\infty}^{\infty} \{f(x)\}^2\, dx = \frac{1}{2\sigma\sqrt{\pi}}$.

[*Hint* : Here $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}}$, $-\infty < x < \infty$.

So $\int_{-\infty}^{\infty} \{f(x)\}^2\, dx = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2}}\, dx$

where we note that $\int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2}}\, dx$ is absolutely convergent.

Now $\int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2}}\, dx = 2\int_{0}^{\infty} e^{-\frac{x^2}{\sigma^2}}\, dx = 2\sigma.\ \tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \sqrt{\pi}\ \sigma$.

Hence $\int_{-\infty}^{\infty} \{f(x)\}^2\, dx = \frac{1}{2\sqrt{\pi}\,\sigma}$.]

2. If X normally distributed with mean 11 and standard deviation 1·5, then find the number a such that

(*i*) $P(X>a)=0.3$, (*ii*) $P(X>a)=0.09$.

[Given Φ (0·53) = 0·7, Φ (0·54) = 0·91.]

3. In a normal distribution 31% items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution. [*I. S. I.* (*Cal.*) '45]

[Given Φ (0·5) = 0·69, Φ (1·40) = 0·92.]

4. If X_k's are mutually independent normal (m_k, σ_k) variates $(k=1, 2, \cdots, n)$, then find the distribution of $\sum_{k=1}^{n} \frac{(X_k-m_k)^2}{\sigma_k^2}$.

[*C. H.* (*Math.*) '71, '73]

[*Hint* : $\frac{X_1-m_1}{\sigma_1}, \frac{X_2-m_2}{\sigma_2}, \ldots\ldots, \frac{X_n-m_n}{\sigma_n}$ are mutually independent standard normal variates. Then by the Theorem 9.2.2,

$\sum_{k=1}^{n}\left(\frac{X_k-m_k}{\sigma_k}\right)^2$ has χ^2 (n) distribution.]

5. The Cartesian co-ordinates (X, Y, Z) of a random point moving in space are mutually independent, each of which is normal (0, 1). Prove that the square of the distance of the point from the origin is χ^2 – distributed. [*C. H.* (*Math.*) '75]

[*Hint* : See Theorem 9.2.2.]

6. If the joint distribution of X and Y be a bivariate normal distribution with parameters $m_x, m_y, \sigma_x, \sigma_y, \rho$, then find the distribution of $\frac{1}{1-\rho^2}\left\{\frac{(X-m_x)^2}{\sigma_x^2} - 2\rho\frac{(X-m_x)(Y-m_y)}{\sigma_x\sigma_y} + \frac{(Y-m_y)^2}{\sigma_y^2}\right\}$.

[*Hint* : We know that $\frac{X-m_x}{\sigma_x}$, $\frac{1}{\sqrt{1-\rho^2}}\left(\frac{Y-m_y}{\sigma_y} - \rho\frac{X-m_x}{\sigma_x}\right)$ are independent standard normal variates see Examples VIII, 16].

Then $\frac{(X-m_x)^2}{\sigma_x^2} + \frac{1}{1-\rho^2}\left(\frac{Y-m_y}{\sigma_y} - \rho\frac{X-m_x}{\sigma_x}\right)^2$ is a χ^2 (2) variate.]

7. If X, Y, Z are mutually independent standard normal variates, find then value of

$P(X^2+Y^2+Z^2-XY-YZ-ZX > 4.8285)$.

[*Hint* : $P(X^2+Y^2+Z^2-XY-YZ-ZX > 4.8285)$

$= P\{X^2+Y^2+Z^2 - \left(\frac{X+Y+Z}{\sqrt{3}}\right)^2 > 3.219\}$

$= P(\chi^2 > 3.219)$,

where by the Theorem 9.2.4.

$\chi^2 = X^2+Y^2+Z^2 - \left(\frac{X+Y+Z}{\sqrt{3}}\right)^2$

has χ^2 – distribution with $3-1=2$ degrees of freedom.

Now from table II (appendix) we find that

$$\chi^2_{\cdot 0\cdot 20, 2} = 3\cdot 219,$$

Hence,the required probability is 0·20.]

8. Let X_1, X_2, X_3, X_4 be mutually independent standard normal variates. Find the value of

$$P\left\{\frac{X_1+X_2+X_3}{\sqrt{\frac{1}{2}(X_1-X_2)^2+\frac{1}{6}(X_1-2X_2+X_3)^2+X_4^2}} < 1\cdot 638\right\}$$

[Given $t_{0\cdot 10, 3} = 1\cdot 638$.]

[*Hint* : Here $X_1+X_2+X_3$ is a normal $(0, \sqrt{3})$ variate.

Also $\dfrac{X_1-X_2}{\sqrt{2}}$, $\dfrac{X_1-2X_2+X_3}{\sqrt{6}}$, X_4 are standard normal variates. Further it can be shown that

$\dfrac{X_1-X_2}{\sqrt{2}}$, $\dfrac{X_1-2X_2+X_3}{\sqrt{6}}$, X_4 are mutually independent variates. So $\dfrac{(X_1-X_2)^2}{2}+\frac{1}{6}(X_1-2X_2+X_3)^2+X_4^2 = V$ (say) is a $\chi^2(3)$ variate.

$$\text{So } T = \frac{X_1+X_2+X_3}{\sqrt{\frac{1}{2}(X_1-X_2)^2+\frac{1}{6}(X_1-2X_2+X_3)^2+X_4^2}}$$

$$= \frac{\dfrac{X_1+X_2+X_3}{\sqrt{3}}}{\sqrt{\dfrac{V}{3}}}$$

where $\dfrac{X_1+X_2+X_3}{\sqrt{3}}$ is a standard normal variate and V is a $\chi^2(3)$ variate and it can be shown that $\dfrac{X_1+X_2+X_3}{\sqrt{3}}$, V are independent.

So T has t-distribution with 3 degrees of freedom.

Hence the required probability is

$$P(T \leqslant 1\cdot 638) = 1 - P(T > 1\cdot 638) = 1 - 0\cdot 10 = 0\cdot 90.\ ]$$

9. If X is an $F(m, n)$ variate and Y is an $F(n, m)$ variate, then prove that

$$P(X > a) + P\left(Y > \frac{1}{a}\right) = 1 \text{ where } a > 0$$

[*Hint* : $P\left(Y > \dfrac{1}{a}\right) = 1 - P\left(Y < \dfrac{1}{a}\right) = 1 - P\left(\dfrac{1}{Y} > a\right)$.

Now $\dfrac{1}{Y}$ is an $F(m, n)$ variate, so that $\dfrac{1}{Y}$ can be identified with X.

$$\text{Hence, } P(X > a) + P\left(Y > \frac{1}{a}\right) = P(X > a) + 1 - P(X > a)$$

$$= P(X > a) + 1 - P(X > a) = 1.\ ]$$

10. If $X_1, X_2, \ldots, X_n$ are mutually independent normal $(0, \sigma)$ variates, then find the distribution of the sum of their squares.

[C. H. (Math.) '64, '66, '68]

Hint : $\dfrac{X_1}{\sigma}, \dfrac{X_2}{\sigma}, \ldots, \dfrac{X_n}{\sigma}$ are mutually independent standard normal variates. Then $\dfrac{X_1^2+X_2^2+\cdots+X_n^2}{\sigma^2}$ is a $\chi^2(n)$ variate.

Let $Y = \dfrac{X}{\sigma^2}$, where $X = X_1^2+X_2^2+\cdots+X_n^2$

Then we get, $X = \sigma^2 Y$ where Y is a $\chi^2(n)$ variate. In real variables we have $x = \sigma^2 y$.

Then $\dfrac{dx}{dy} = \sigma^2 > 0$ for all y.

So the probability density function $f_X(x)$ of X is given by

$$f_X(x) = \frac{1}{\sigma^2} f_Y(y)$$

$$\text{or, } f_X(x) = \frac{1}{\sigma^2} \frac{e^{-\frac{y}{2}}\left(\frac{y}{2}\right)^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n}{2}\right)} \quad \text{if } y > 0$$

$$\text{or, } f_X(x) = \frac{e^{-\frac{x}{2\sigma^2}}\left(\frac{x}{2\sigma^2}\right)^{\frac{n}{2}-1}}{2\sigma^2\Gamma\left(\frac{n}{2}\right)} \quad \text{if } x > 0$$

which gives the distribution of $X = X_1^2+X_2^2+\cdots+X_n^2$.]

11. If X, $X+Y$ are respectively $\chi^2(m)$ and $\chi^2(m+n)$ variates and X, Y are independent, then show that Y is a $\chi^2(n)$ variate.

[C. H. (Math.) '69]

[*Hint* : Characteristic function $\phi_{X+Y}(t)$ of $X+Y$ is given by

$$\phi_{X+Y}(t)=E(e^{itX}\cdot e^{itY})=E(e^{itX})E(e^{itY})$$

or, $\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t)$.

Here $\phi_{X+Y}(t)=(1-2it)^{-\frac{m+n}{2}}$

$$\phi_X(t)=(1-2it)^{-\frac{m}{2}}$$

So $\phi_Y(t)=(1-2it)^{-\frac{n}{2}}$ which is the characteristic function of a $\chi^2(n)$ variate.]

12. If χ_1^2, χ_2^2 are independent χ^2-variates with m and n degrees of freedom respectively, then find the distribution of

$$\frac{\chi_1^2}{\chi_2^2}.$$

[*C. H.* (*Math.*) '69]

[*Hint* : Here $\dfrac{\frac{\chi_1^2}{m}}{\frac{\chi_2^2}{n}}$ is an $F(m, n)$ variate. So writing $X=\dfrac{\frac{\chi_1^2}{m}}{\frac{\chi_2^2}{n}}$ and

and $Y=\dfrac{\chi_1^2}{\chi_2^2}$ we find $Y=\dfrac{m}{n}X$, where X is an $F(m, n)$ variate. Then the probability density function $f_Y(y)$ of Y will be given by

$$f_Y(y)=\left|\frac{dx}{dy}\right|f_X(x) \text{ where}$$

$$f_X(x)=\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}x^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\ \frac{n}{2}\right)\left(mx+n\right)^{\frac{m+n}{2}}} \text{ if } x>0.$$

So $$f_Y(y)=\frac{n}{m}\cdot\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}\left(\frac{ny}{m}\right)^{\frac{m}{2}-1}}{B\left(\frac{m}{2},\ \frac{n}{2}\right)\left(ny+n\right)^{\frac{m+n}{2}}} \text{ if } y>0$$

which determines the distribution of $Y=\dfrac{\chi_1^2}{\chi_2^2}$.]

13. Show that the variance of the t-distribution with n degrees of freedom exists if and only if $n>2$ and that its value is $\dfrac{n}{n-2}$

[*Hint* : See § 9·3.] [*C. H.* (*Math.*) '70]

14. If the Cartesian co-ordinates (X, Y, Z) of a random point in space be mutually independent standard normal variates, then show that the square of the distance of the origin from the foot of the perpendicular from the random point to the plane $lx+my+nz=0$, $(l^2+m^2+n^2=1)$, has χ^2 distribution with 2 degrees of freedom

15. If X_1, X_2, X_3, X_4, X_5 are mutually independent normal (0, 4) variates, then find the value of

$$P\left(\frac{X_1^2+X_2^2+X_3^2}{16}>3-\frac{X_4^2+X_5^2}{16}\right).\ [\text{Given } \chi^2_{0\cdot70,5}=3.]$$

Answers

2. (*i*) 11·795, (*ii*) 11·690.
3. 50, 10.
7. 0·20.
15. 0·70.

CHAPTER X

CONVERGENCE OF A SEQUENCE OF RANDOM VARIABLES AND LIMIT THEOREMS

10.1. Introduction.

If for each positive integer n a random variable X_n is defined on a given event space S (same for each n) with respect to a given class of events Δ and a probability function $P : \Delta \to R$, then we say that $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of random variables and as in analysis we denote the sequence by $\{X_n\}$.

From practical point of view the discussion of a random variable X will be highly significant if it is known that there exists a real constant a for which $P(|X-a|<\varepsilon) \approx 1$, where $\varepsilon\,(>0)$ is sufficiently small, that is, it is nearly certain that values of X lie in a very small neighbourhood of a.

For a sequence of random variables $\{X_n\}$, each X_n may not have the above property but it may happen that the aforesaid property (with respect to a real constant a) becomes more and more distinguished as n gradually increases and the question of existence of such a real constant 'a' will be answered by the concept of convergence in probability of the sequence $\{X_n\}$.

Again the sequence $\{X_n\}$ may be such that as n gradually increases the distribution function $F_n(x)$ of X_n may more and more resemble to the distribution function of a particular random variable and the question of existence of such a distribution function is related to the concept 'convergence in distribution' of the sequence $\{X_n\}$.

Besides the above mentioned two modes of convergence, there are other modes of convergence of the sequence $\{X_n\}$. In this chapter we shall limit our discussion to three types of convergence of a sequence of random variables, namely,

(*i*) Convergence in probability,
(*ii*) Convergence in mean square,
(*iii*) Convergence in distribution.

We will conclude the chapter discussing some fundamental limit theorems related to the modes of convergence (*i*) and (*iii*) — Law of Large Numbers (Weak Law), Central limit theorem, De Moivre-Laplace limit theorem and limit theorem for characteristic functions.

In the next section we shall prove some important inequalities of which the most important is 'Tchebycheff's Inequality' which will be used in proving many theorems on convergence in probability.

10.2. Some Fundamental Inequalities.

A. Theorem 10.2.1. Tchebycheff's Inequality.

If X be any random variable having finite variance σ^2 (and hence having finite mean m), then for any $\varepsilon>0$,

$$P(|X-m|\ge\varepsilon)\le\frac{\sigma^2}{\varepsilon^2}. \tag{10.2.1}$$

Proof: Case I. Let X be a discrete random variable.

Then
$$P(|X-m|\ge\varepsilon)=\sum_{|x_i-m|\ge\varepsilon} f_i, \tag{10.2.2}$$

where $P(X=x_i)=f_i$, x_i being a point of the spectrum of X. Now for each x_i satisfying $|x_i-m|\ge\varepsilon$, we have

$$\frac{(x_i-m)^2}{\varepsilon^2}\ge 1,$$

so
$$f_i\le\frac{(x_i-m)^2}{\varepsilon^2}f_i, \quad \text{since } f_i\ge 0$$

for all x_i satisfying $|x_i-m|\ge\varepsilon$.

Therefore
$$\sum_{|x_i-m|\ge\varepsilon} f_i \le \frac{1}{\varepsilon^2}\sum_{|x_i-m|\ge\varepsilon}(x_i-m)^2 f_i$$
$$\le\frac{1}{\varepsilon^2}\sum_{i=-\infty}^{\infty}(x_i-m)^2 f_i,$$

since $(x_i-m)^2 f_i\ge 0$ for every point x_i of the spectrum.

But
$$\sum_{i=-\infty}^{\infty}(x_i-m)^2 f_i = \text{variance of } X=\sigma^2.$$

Therefore
$$\sum_{|x_i-m|\ge\varepsilon} f_i\le\frac{\sigma^2}{\varepsilon^2}.$$

Hence from (10.2.2), we get $P(|X-m|\ge\varepsilon)\le\dfrac{\sigma^2}{\varepsilon^2}$.

Case II. Let X be a continuous random variable.

Here
$$P(|X-m|\ge\varepsilon)=\int_{|x-m|\ge\varepsilon} f(x)\,dx, \tag{10.2.3}$$

where $f(x)$ is the probability density function of X.

Now $|x-m|\geq\varepsilon \Leftrightarrow x\in(-\infty, m-\varepsilon]\cup[m+\varepsilon, \infty)$.

Therefore $P(|X-m|\geq\varepsilon)=\int_{-\infty}^{m-\varepsilon} f(x)\,dx+\int_{m+\varepsilon}^{\infty} f(x)\,dx.$ (10.2.4)

Now $\frac{(x-m)^2}{\varepsilon^2}\geq 1$ whenever $|x-m|\geq\varepsilon$

so $\frac{f(x)(x-m)^2}{\varepsilon^2}\geq f(x)$ for all $x\in(-\infty, m-\varepsilon]\cup[m+\varepsilon,\infty)$,

since $f(x)\geq 0$ for all x.

Therefore $$\int_{-\infty}^{m-\varepsilon} f(x)\,dx\leq\int_{-\infty}^{m-\varepsilon}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx$$

and $$\int_{m+\varepsilon}^{\infty} f(x)\,dx\leq\int_{m+\varepsilon}^{\infty}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx.$$

Hence from (10.2.4), we get

$$P(|X-m|\geq\varepsilon)\leq\int_{-\infty}^{m-\varepsilon}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx+\int_{m+\varepsilon}^{\infty}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx$$

$$\leq\int_{-\infty}^{m-\varepsilon}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx+\int_{m-\varepsilon}^{m+\varepsilon}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx+\int_{m+\varepsilon}^{\infty}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx,$$

since $\int_{m-\varepsilon}^{m+\varepsilon}\frac{(x-m)^2}{\varepsilon^2}f(x)\,dx\geq 0$.

Therefore $P(|X-m|\geq\varepsilon)\leq\frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-m)^2 f(x)\,dx=\frac{\sigma^2}{\varepsilon^2}$.

Hence the inequality.

Cor. If τ be any positive number, taking $\varepsilon=\sigma\tau>0$, we get, from (10.2.1),

$$P(|X-m|\geq\sigma\tau)\leq\frac{\sigma^2}{\sigma^2\tau^2}=\tau^{-2}$$

or, $$1-P(|X-m|<\sigma\tau)\leq\tau^{-2}$$

or, $$P(|X-m|<\sigma\tau)\geq 1-\tau^{-2}. \tag{10.2.5}$$

Remark. Ordinarily, the distribution of a random variable is required to obtain any probability connected to the random variable. But here in Tchebycheff's inequality, we get an upper bound for the probability that the deviation of X from its mean is at least ε units in terms of variance of X and ε. No assumption on the distribution of X is made other than that it has finite variance. In other words Tchebycheff's inequality indicates that irrespective of the shape of the density curve $P(m-\varepsilon<X<m+\varepsilon)>1-\frac{\sigma^2}{\varepsilon^2}$, that is, the probability that X takes values in the interval $(m-\varepsilon, m+\varepsilon)$ centred at m is close to 1, provided $\frac{\sigma}{\varepsilon}$ is sufficiently small.

B. Theorem 10.2.2. *Let $g:R\to R$ be a continuous function such that $g(x)\geq 0$ for all $x\in R$. If X be a discrete random variable and if $E\{g(X)\}$ exists, then for any $\varepsilon>0$,*

$$P\{g(X)\geq\varepsilon\}\leq\frac{E\{g(X)\}}{\varepsilon} \tag{10.2.6}$$

Proof: Here X is a discrete random variable.

Then $E\{g(X)\}=\sum_{g(x_i)\geq\varepsilon} g(x_i)f_i+\sum_{g(x_i)<\varepsilon} g(x_i)f_i$,

where $P(X=x_i)=f_i$, x_i being a point of the spectrum of X. Since $g(x_i)\geq 0$, $f_i\geq 0$ for all x_i, we get

$$E\{g(X)\}\geq\sum_{g(x_i)\geq\varepsilon} g(x_i)f_i\geq\varepsilon\sum_{g(x_i)\geq\varepsilon} f_i=\varepsilon P\{g(X)\geq\varepsilon\}.$$

Therefore, $P\{g(X)\geq\varepsilon\}\leq\frac{E\{g(X)\}}{\varepsilon}$, since $\varepsilon>0$

Note. The inequality (10.2.6) can also be proved for a continuous variate if we use the concept of Borel measurable function and Lebesgue- Stieltje's integrals (which is beyond the scope of this treatise).

Cor.1. In particular if we take the non-negative function $g:R\to R$ such that $g(x)=(x-m)^2$, then replacing ε by ε^2, we get from (10.2.6),

$$P\{(X-m)^2\geq\varepsilon^2\}\leq\frac{E\{(X-m)^2\}}{\varepsilon^2}.$$

Since the event '$(X-m)^2\geq\varepsilon^2$' implies and is implied by the event '$|X-m|\geq\varepsilon$', we get Tchebycheff's inequality

$$P(|X-m|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}.$$

Cor. 2. Considering the non-negative function $g : R \to R$ such that $g(X) = |X|^r$, where $r > 0$ and replacing ε by ε^r $(\varepsilon > 0)$ we get, from (10.2.6),

$$P(|X|^r \geq \varepsilon^r) \leq \frac{E(|X|^r)}{\varepsilon^r}$$

or,
$$P(|X| \geq \varepsilon) \leq \frac{E(|X|^r)}{\varepsilon^r}. \tag{10.2.7}$$

C. Theorem 10.2.3. *Let $g : R \to R$ be a non-decreasing function such that $g(x) > 0$ for all $x \in R$ and $E(X) = m$, where X is a random variable. If $E\{g(|X-m|)\}$ exists, then for any $\varepsilon > 0$,*

$$P(|X-m| \geq \varepsilon) \leq \frac{E\{g(|X-m|)\}}{g(\varepsilon)}. \tag{10.2.8}$$

Proof : Case I. Let X be a discrete random variable. Then

$$E\{g(|X-m|)\} = \sum_{i=-\infty}^{\infty} g(|x_i - m|) f_i,$$

where $P(X = x_i) = f_i$, x_i being a point of the spectrum of X.

Therefore, $E\{g(|X-m|)\}$

$$= \sum_{|x_i - m| \geq \varepsilon} g(|x_i - m|) f_i + \sum_{|x_i - m| < \varepsilon} g(|x_i - m|) f_i$$

$$\geq \sum_{|x_i - m| \geq \varepsilon} g(|x_i - m|) f_i,$$

since $f_i \geq 0$ for all i and $g(x) > 0$ for all x.

Since $g(x)$ is non-decreasing,

$$|x_i - m| \geq \varepsilon \text{ implies } g(|x_i - m|) \geq g(\varepsilon).$$

Therefore, $E\{g(|X-m|)\} \geq g(\varepsilon) \sum_{|x_i - m| \geq \varepsilon} f_i$

$$= g(\varepsilon) P(|X-m| \geq \varepsilon).$$

Therefore, $P(|X-m| \geq \varepsilon) \leq \dfrac{E\{g(|X-m|)\}}{g(\varepsilon)}$ since $g(\varepsilon) > 0$.

Case II. Let X be a continuous random variable with $f(x)$ as its probability density function. Then

$$E\{g(|X-m|)\} = \int_{-\infty}^{\infty} g(|x-m|)\, f(x)\, dx$$

$$= \int_{|x-m| \geq \varepsilon} g(|x-m|) f(x)\, dx + \int_{|x-m| < \varepsilon} g(|x-m|)\, f(x)\, dx$$

$$\geq \int_{|x-m| \geq \varepsilon} g(|x-m|)\, f(x)\, dx, \text{ since } f(x) \geq 0 \text{ and } g(x) > 0.$$

Now since $g(x)$ is non-decreasing,

$$|x-m| \geq \varepsilon \text{ implies } g(|x-m|) \geq g(\varepsilon).$$

Therefore, $E\{g(|X-m|)\} \geq g(\varepsilon) \displaystyle\int_{|x-m| \geq \varepsilon} f(x)\, dx$

$$= P(|X-m| \geq \varepsilon)\, g(\varepsilon)$$

Hence, $P(|X-m| \geq \varepsilon) \leq \dfrac{E\{g(|X-m|)\}}{g(\varepsilon)}$,

since $g(\varepsilon) > 0$.

D. Theorem 10.2.4. (*Generalisation of Tchebycheff's Inequality*). *If X possesses a finite second order moment about c, where c is any fixed number, then, for any $\varepsilon > 0$.*

$$P(|X-c| \geq \varepsilon) \leq \frac{E\{(X-c)^2\}}{\varepsilon^2}. \tag{10.2.9}$$

Proof : Case I.

Let X be a discrete random variable.

Then
$$P(|X-c| \geq \varepsilon) = \sum_{|x_i - c| \geq \varepsilon} f_i, \tag{10.2.10}$$

where $P(X = x_i) = f_i$, x_i being a point of the spectrum of X. Now for each x_i satisfying $|x_i - c| \geq \varepsilon$, $\dfrac{(x_i - c)^2}{\varepsilon^2} \geq 1$ and so $f_i \leq \dfrac{(x_i - c)^2}{\varepsilon^2} f_i$, since $f_i \geq 0$, for all x_i satisfying $|x_i - c| \geq \varepsilon$.

Therefore, $\sum_{|x_i-c|\geq\varepsilon} f_i \leq \frac{1}{\varepsilon^2} \sum_{|x_i-c|\geq\varepsilon} (x_i-c)^2 f_i$

$$\leq \frac{1}{\varepsilon^2} \sum_{i=-\infty}^{\infty} (x_i-c)^2 f_i, \tag{10.2.11}$$

since $(x_i-c)^2 f_i \geq 0$ for every point x_i of the spectrum of X.

But $\sum_{i=-\infty}^{\infty} (x_i-c)^2 f_i = E\{(X-c)^2\}$.

Hence, from (10.2.11), $P(|X-c|\geq\varepsilon) \leq \dfrac{E\{(X-c)^2\}}{\varepsilon^2}$.

Case II. Let X be a continuous variate.

Here $P(|X-c|\geq\varepsilon) = \int_{|x-c|\geq\varepsilon} f(x)\,dx$ (10.2.12)

where $f(x)$ is the probability density function of X.

Now $|x-c|\geq\varepsilon$ implies and is implied by $x\in(-\infty, c-\varepsilon] \cup [c+\varepsilon, \infty)$.

Therefore, $P(|X-c|\geq\varepsilon) = \int_{-\infty}^{c-\varepsilon} f(x)\,dx + \int_{c+\varepsilon}^{\infty} f(x)\,dx$. (10.2.13)

Now $\dfrac{(x-c)^2}{\varepsilon^2} \geq 1$ whenever $|x-c|\geq\varepsilon$.

So $\dfrac{f(x)(x-c)^2}{\varepsilon^2} \geq f(x)$

for all $x\in(-\infty, c-\varepsilon]\cup[c+\varepsilon,\infty)$, since $f(x)\geq 0$ for all x.

Therefore, $\int_{-\infty}^{c-\varepsilon} f(x)\,dx \leq \int_{-\infty}^{c-\varepsilon} \dfrac{(x-c)^2}{\varepsilon^2} f(x)\,dx$

and $\int_{c+\varepsilon}^{\infty} f(x)\,dx \leq \int_{c+\varepsilon}^{\infty} \dfrac{(x-c)^2}{\varepsilon^2} f(x)\,dx$.

Hence from (10.2.13), we get

$$P(|X-c|\geq\varepsilon) \leq \int_{-\infty}^{c-\varepsilon} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx + \int_{c+\varepsilon}^{\infty} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx$$

$$\leq \int_{-\infty}^{c-\varepsilon} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx + \int_{c-\varepsilon}^{c+\varepsilon} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx + \int_{c+\varepsilon}^{\infty} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx.$$

since $\int_{c-\varepsilon}^{c+\varepsilon} \dfrac{(x-c)^2}{\varepsilon^2} f(x)\,dx \geq 0$.

Therefore, $P(|X-c|\geq\varepsilon) \leq \dfrac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x-c)^2 f(x)\,dx$

$$= \frac{E\{(X-c)^2\}}{\varepsilon^2}.$$

Hence the inequality.

10.3. Different Types of Convergence of a Sequence of Random Variables.

A. Convergence in Probability.

Let $\{X_n\}$ be a sequence of random variables, where each X_n is defined on the same event space S with respect to a given class of subsets (of S) as the class Δ of events and a given probability function $P:\Delta\to R$. The sequence $\{X_n\}$ is said to be convergent in probability to a real constant a, if for any $\varepsilon>0$,

$$\lim_{n\to\infty} P(|X_n-a|\geq\varepsilon)=0 \tag{10.3.1}$$

or equivalently,

$$\lim_{n\to\infty} P(|X_n-a|<\varepsilon)=1 \tag{10.3.2}$$

and we write $X_n \xrightarrow[\text{in p}]{} a$ as $n\to\infty$.

The sequence $\{X_n\}$ is said to be convergent in probability to a random variable X (defined on the same event space S) if for any $\varepsilon>0$,

$$\lim_{n\to\infty} P(|X_n-X|\geq\varepsilon)=0 \tag{10.3.3}$$

or equivalently,

$$\lim_{n\to\infty} P(|X_n-X|<\varepsilon)=1 \tag{10.3.4}$$

and we write $X_n \xrightarrow[\text{in p}]{} X$ as $n\to\infty$,

where "$|X_n-X|<\varepsilon$" denotes the event

$$\{\omega : \omega\in S \text{ and } |X_n(\omega)-X(\omega)|<\varepsilon\}.$$

A single real constant a can be regarded as the only element of the spectrum of a random variable, say X_0, defined on S and so

$$X_0(\omega)=a \text{ for all } \omega\in S.$$

Then the event " $|X_n - a| < \varepsilon$ " can also be expressed as $\{\omega : \omega \in S$ and $|X_n(\omega) - X_0(\omega)| < \varepsilon\}$ and so (10.3.1) and (10.3.2) can respectively be expressed as

$$\lim_{n\to\infty} P(|X_n - X_0| \geq \varepsilon) = 0$$

and $$\lim_{n\to\infty} P(|X_n - X_0| < \varepsilon) = 1.$$

Hence, the definition given in (10.3.1) or (10.3.2) can be regarded as a particular case of the definition (10.3.3) or (10.3.4) so that convergence in probability of a sequence $\{X_n\}$ to a real constant a can also be regarded as convergence in probability to a random variable X_0 whose spectrum contains the only number a.

From now, in mentioning a sequence $\{X_n\}$, the event space S, the class Δ of events and the probability function $P : \Delta \to R$ will not be mentioned and these are understood to be given.

B. Convergence in Mean Square.

A sequence of random variables $\{X_n\}$ is said to be convergent in mean square to a random variable X (defined in the same event space S) if

$$\lim_{n\to\infty} E\{(X_n - X)^2\} = 0, \quad \ldots \quad (10.3.5)$$

provided (*i*) $E(X^2)$ exists,

(*ii*) $E(X_n^2)$ exists for all n,

and we write $X_n \xrightarrow{2} X$ as $n \to \infty$. The random variable X is called the limit in mean square of $\{X_n\}$.

C. Convergence in Distribution.

Let $\{X_n\}$ be a sequence of random variables, where $F_n(x)$ is the distribution function of X_n for $n = 1, 2, 3, \ldots$ If there exists a random variable X whose distribution function is $F(x)$ such that $\lim_{n\to\infty} F_n(x) = F(x)$ at every point of continuity x of $F(x)$, then $\{X_n\}$ is said to be convergent in distribution or convergent in law to X and we write

$$X_n \longrightarrow X \quad \text{as } n \to \infty$$

or, $$X_n \xrightarrow{L} X \quad \text{as } n \to \infty.$$

Remarks. (1) In real analysis, a sequence $\{x_n\}$ is said to converge to a real number l, if corresponding to any $\varepsilon > 0$, there exists a positive integer N, such that $|x_n - l| < \varepsilon$ for all $n > N$. In the above definition of convergence in probability of a sequence of random variables $\{X_n\}$ to a constant a, we say nothing about the convergence of the random variable X_n to 'a' in the same sense as we understand in analysis. The definition (10.3.1) does not imply that for every $\varepsilon > 0$, we can find a positive integer N such that

$$|X_n(\omega) - a| < \varepsilon \quad \text{for all} \quad n > N$$

is satisfied for every $\omega \in S$.

In fact convergence in probability implies that the sequence of probabilities $P(|X_n - a| \geq \varepsilon)$ tend to zero as $n \to \infty$. Similar observation holds when a sequence of random variables $\{X_n\}$ converges in probability to a random variable X.

(2) From definition (10.3.1), we observe that if n is very large $P(|X_n - a| < \varepsilon) \approx 1$, that is, the event '$|X_n - a| < \varepsilon$' is nearly certain. Now ε can be chosen very small. Then the event '$|X_n - a| < \varepsilon$' can be interpreted as the event 'values of X_n lie very close to a'. So $X_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$ implies that as n increases $P(|X_n - a| < \varepsilon)$ gradually approaches 1 and we become more and more sure that values of X_n lie in a small neighbourhood of 'a'.

Theorem 10.3.1. If $\underset{n\to\infty}{Lt}\, u_n = l$, then $u_n \xrightarrow[\text{in p}]{} l$ as $n \to \infty$.

Proof: Here for every n, u_n can be regarded as a random variable with one element u_n in the corresponding spectrum. Let $\varepsilon > 0$ be any real number. Since $\underset{n\to\infty}{Lt}\, u_n = l$, there exists a positive integer m such that

$$|u_n - l| < \varepsilon \text{ for all } n > m.$$

Now the event $\{\omega : \omega \in S$ and $|u_n(\omega) - l| < \varepsilon\}$ is the same as the event '$|u_n - l| < \varepsilon$' which is here the certain event S for all $n > m$.

So
$$\begin{aligned} P\{\omega : \omega \in S \text{ and } |u_n(\omega) - l| < \varepsilon\} &= P(|u_n - l| < \varepsilon) \\ &= P(S) \text{ for all } n > m \\ &= 1 \quad \text{for all } n > m. \end{aligned}$$

Hence $\underset{n\to\infty}{Lt}\, P\{\omega : \omega \in S \text{ and } |u_n(\omega) - l| < \varepsilon\} = 1$ for every $\varepsilon > 0$.

So $$u_n \xrightarrow[\text{in p}]{} l \quad \text{as } n \to \infty.$$

Theorem 10.3.2. *If* $X_n \xrightarrow{2} X$ *as* $n \to \infty$, *then* $X_n \xrightarrow[\text{in } p]{} X$ *as* $n \to \infty$, *that is, convergence in mean square implies convergence in probability.*

Proof: Here $E\{(X_n - X)^2\}$ exists for all n and $\lim_{n\to\infty} E\{(X_n - X)^2\} = 0$.

Then by generalisation of Tchebycheff's inequality, for any $\varepsilon > 0$, we have $P(|X_n - X| \geq \varepsilon) \leq \dfrac{E\{(X_n - X)^2\}}{\varepsilon^2}$.

Then we have

$$0 \leq P(|X_n - X| \geq \varepsilon) \leq \frac{E\{(X_n - X)^2\}}{\varepsilon^2} \text{ for all } n,$$

where $\lim_{n\to\infty} \dfrac{E\{(X_n - X)^2\}}{\varepsilon^2} = 0$.

Hence, $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$.

Therefore, $X_n \xrightarrow[\text{in } p]{} X$ as $n \to \infty$.

Remark. The converse of Theorem 10.3.2 is not true. We consider the sequence of random variables $\{X_n\}$ defined by

$X_n = 0, n$ with $P(X_n = 0) = 1 - \dfrac{1}{n^2}$ and $P(X_n = n) = \dfrac{1}{n^2}$.

Then $E(X_n^2) = 0.\left(1 - \dfrac{1}{n^2}\right) + n^2 . \dfrac{1}{n^2} = 1 \neq 0$.

Therefore, $E(X_n^2)$ does not tend to zero as $n \to \infty$, which implies that $\{X_n\}$ does not converge in mean square to 0. Again for any $\varepsilon > 0$,

$$P(|X_n| \geq \varepsilon) = P(X_n = n), \text{ if } 0 < \varepsilon \leq n$$
$$= 0 \quad \text{if } \varepsilon > n$$

or,
$$P(|X_n| \geq \varepsilon) = \frac{1}{n^2} \quad \text{if } 0 < \varepsilon \leq n$$
$$= 0 \quad \text{if } \varepsilon > n$$

Therefore, $P(|X_n| \geq \varepsilon) \to 0$ as $n \to \infty$ and hence

$$X_n \xrightarrow[\text{in } p]{} 0 \text{ as } n \to \infty.$$

Theorem 10.3.3. *Let* $\{X_n\}$ *be a sequence of random variables such that* $X_n \xrightarrow{2} X$ *as* $n \to \infty$. *Then* $E(X_n) \to E(X)$ *and* $E(X_n^2) \to E(X^2)$ *as* $n \to \infty$.

Proof: Since $X_n \xrightarrow{2} X$ as $n \to \infty$, we have

$$E\{(X_n - X)^2\} \longrightarrow 0 \text{ as } n \to \infty. \quad (10.3.6)$$

Now $E(X^2) - \{E(X)\}^2 = E\{(X - m)^2\} \geq 0$, where $m = E(X)$.

So for any random variable X, $E(X^2) \geq \{E(X)\}^2$ if $E(X^2)$ exists. Then for the random variable $|X|$ we get

$$E(|X|^2) \geq \{E(|X|)\}^2$$

that is, $E(|X|) \leq \sqrt{E(X^2)}$ (10.3.7)

Now $|E(X_n) - E(X)| = |E(X_n - X)| \leq E(|X_n - X|)$, by the property VI, p. 366

$\leq \sqrt{E\{(X_n - X)^2\}}$ by (10.3.7).

Thus $0 \leq |E(X_n) - E(X)| \leq \sqrt{E\{(X_n - X)^2\}}$

where $\underset{n\to\infty}{Lt}\, E\{(X_n - X)^2\} = 0$.

Therefore, $\underset{n\to\infty}{Lt}\, |E(X_n) - E(X)| = 0$,

or, $E(X_n) \to E(X)$ as $n \to \infty$.

Again, $X_n^2 = (X_n - X)^2 + X^2 + 2(X_n - X)X$.

Therefore, $E(X_n^2) = E\{(X_n - X)^2\} + E(X^2) + 2E\{(X_n - X)X\}$. (10.3.8)

Now by Schwarz's inequality

$$|E\{X(X_n - X)\}| \leq \sqrt{E(X^2)\, E\{(X_n - X)^2\}}.$$

Then by (10.3.6), $E\{X(X_n - X)\} \to 0$ as $n \to \infty$.

Therefore, from (10.3.8), $E(X_n^2) \to E(X^2)$ as $n \to \infty$.

10.4. Some Results for Convergence in Probability.

Theorem 10.4.1. If $X_n \xrightarrow[\text{in } p]{} a$, $Y_n \xrightarrow[\text{in } p]{} b$ as $n \to \infty$, then

(i) $X_n - a \xrightarrow[\text{in } p]{} 0$ as $n \to \infty$,

(ii) $cX_n \to ca$ as $n \to \infty$, c being constant,

(iii) $X_n \pm Y_n \xrightarrow[\text{in } p]{} a \pm b$ as $n \to \infty$,

(iv) $X_n^2 \xrightarrow[\text{in } p]{} a^2$ as $n \to \infty$,

(v) $X_n Y_n \xrightarrow[\text{in } p]{} ab$ as $n \to \infty$,

(vi) $\dfrac{X_n}{Y_n} \xrightarrow[\text{in } p]{} \dfrac{a}{b}$ as $n \to \infty$, provided $b \neq 0$.

Proof: (i) Put $Z_n = X_n - a$. For any $\varepsilon > 0$,

$$\lim_{n\to\infty} P|(Z_n - 0| < \varepsilon) = \underset{n\to\infty}{Lt}\, P(|X_n - a| < \varepsilon) = 1$$

if $X_n \xrightarrow[\text{in } p]{} a$ as $n \to \infty$.

Therefore, $X_n \xrightarrow[\text{in p}]{} a$ implies $Z_n \xrightarrow[\text{in p}]{} 0$, that is, $X_n - a \xrightarrow[\text{in p}]{} 0$, as $n \to \infty$.

(ii) If $c = 0$, for any $\varepsilon > 0$,

$$P(|cX_n - ca| < \varepsilon) = P(|0X_n - 0a| < \varepsilon).$$

Now $|0X_n - 0a| < \varepsilon$ is satisfied for all $\omega \in S$, since $|0X_n(\omega) - 0a| = 0 < \varepsilon$ for all $\omega \in S$.

Therefore, $|0X_n - 0a| < \varepsilon = S$, and hence $P(|0X_n - 0a| < \varepsilon) = 1$ for all n.

Therefore, $\underset{n\to\infty}{Lt}\ P(|0X_n - 0a| < \varepsilon) = 1$

Therefore, $cX_n \xrightarrow[\text{in p}]{} ca$ as $n \to \infty$ if $c = 0$.

If $c \neq 0$, for any $\varepsilon > 0$,

$$P(|cX_n - ca| < \varepsilon) = P\left(|X_n - a| < \frac{\varepsilon}{|c|}\right) \longrightarrow 1 \text{ as } n \to \infty.$$

whenever $X_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$. This again implies that

$$cX_n \xrightarrow[\text{in p}]{} ca \text{ as } n \to \infty.$$

Thus $cX_n \xrightarrow[\text{in p}]{} ca$ as $n \to \infty$ for any constant c whenever $X_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$.

(iii) Let ε be any positive number.

Since $X_n \xrightarrow[\text{in p}]{} a$, $Y_n \xrightarrow[\text{in p}]{} b$ as $n \to \infty$, we have

$$\lim_{n\to\infty} P\left(|X_n - a| \geq \frac{\varepsilon}{2}\right) = 0 \tag{10.4.1}$$

and

$$\lim_{n\to\infty} P\left(|Y_n - b| \geq \frac{\varepsilon}{2}\right) = 0. \tag{10.4.2}$$

Let A_n, B_n and C_n denote respectively the events '$|X_n - a| < \frac{\varepsilon}{2}$', '$|Y_n - b| < \frac{\varepsilon}{2}$' and '$|(X_n + Y_n) - (a+b)| < \varepsilon$'.

If A_n and B_n happen simultaneously, then

$|(X_n + Y_n) - (a+b)| = |(X_n - a) + (Y_n - b)| \leq |X_n - a| + |Y_n - b| < \varepsilon.$

Thus $A_n B_n$ implies C_n, that is, $\omega \in A_n B_n \Rightarrow \omega \in C_n$.

Therefore, $A_n B_n \subseteq C_n$

or, $\overline{A_n B_n} \supseteq \overline{C}_n$, that is, $\overline{C}_n \subseteq \overline{A_n B_n} = \overline{A}_n + \overline{B}_n$.

Therefore, $P(\overline{C}_n) \leq P(\overline{A}_n + \overline{B}_n) = P(\overline{A}_n) + P(\overline{B}_n) - P(\overline{A}_n \overline{B}_n)$

$$\leq P(\overline{A}_n) + P(\overline{B}_n). \tag{10.4.3}$$

Now $\overline{A}_n$, $\overline{B}_n$ and $\overline{C}_n$ represent respectively the events '$|X_n - a| \geq \frac{\varepsilon}{2}$', '$|Y_n - b| \geq \frac{\varepsilon}{2}$' and '$|(X_n + Y_n) - (a+b)| \geq \varepsilon$'.

Hence from (10.4.3),

$$0 \leq P\{|(X_n + Y_n) - (a+b)| \geq \varepsilon\}$$

$$\leq P\left(|X_n - a| \geq \frac{\varepsilon}{2}\right) + P\left(|Y_n - b| \geq \frac{\varepsilon}{2}\right).$$

Therefore, proceeding to the limit as $n \to \infty$ and using (10.4.1), (10.4.2) we get

$$\lim_{n\to\infty} P\{|(X_n + Y_n) - (a+b)| \geq \varepsilon\} = 0 \text{ for every } \varepsilon > 0.$$

In other words $X_n + Y_n \xrightarrow[\text{in p}]{} a + b$ as $n \to \infty$.

Again, let D_n denote the event '$|(X_n - Y_n) - (a-b)| < \varepsilon$'.

Now if A_n and B_n happen simultaneously, then

$$|(X_n - Y_n) - (a-b)| = |(X_n - a) - (Y_n - b)| \leq |X_n - a| + |Y_n - b| < \varepsilon. \tag{10.4.4}$$

Thus $A_n B_n$ implies D_n.

Then proceeding as above,

$$P(\overline{D}_n) \leq P(\overline{A}_n) + P(\overline{B}_n). \tag{10.4.5}$$

Now $\overline{D}_n$ represents the event '$|(X_n - Y_n) - (a-b)| \geq \varepsilon$'.

Hence from (10.4.5),

$$0 \leq P\{|(X_n - Y_n) - (a-b)| \geq \varepsilon\}$$

$$\leq P\left(|X_n - a| \geq \frac{\varepsilon}{2}\right) + P\left(|Y_n - b| \geq \frac{\varepsilon}{2}\right).$$

Hence, proceeding to the limit as $n \to \infty$ and using (10.4.1) and (10.4.2), we get

$$\lim_{n\to\infty} P\{|(X_n - Y_n) - (a-b)| \geq \varepsilon\} = 0,$$

which implies that $X_n - Y_n \xrightarrow[\text{in p}]{} a - b$ as $n \to \infty$.

By the principle of finite induction, the result may be extended to any finite number of sequences of random variables.

(iv) We first note that if $Z_n \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$, then

$$Z_n^2 \xrightarrow[\text{in p}]{} 0 \text{ as } n \to \infty, \text{ since for any } \varepsilon > 0,$$

$P(|Z_n^2| < \varepsilon) = P(|Z_n| < \sqrt{\varepsilon}) \to 1$ as $n \to \infty$, and this implies that $Z_n^2 \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$, whenever $Z_n \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Now $X_n^2 = (X_n - a)^2 + 2a(X_n - a) + a^2 = Z_n^2 + 2a Z_n + a^2$ where $Z_n = X_n - a$. (10.4.6)

Now when $X_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$, $Z_n \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Also $aZ_n \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$, since $Z_n \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$. Also $a^2 \xrightarrow[\text{in p}]{} a^2$ as $n \to \infty$.

Further it has been proved that $Z_n^2 \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Hence, from (10.4.6), using (iii) (for finite number of sequences of random variables),

$$X_n^2 \xrightarrow[\text{in p}]{} a^2 \text{ as } n \to \infty.$$

(v) We have $X_n Y_n = \frac{1}{4}\{(X_n + Y_n)^2 - (X_n - Y_n)^2\}$

$$= \frac{1}{4}(Z_n^2 - T_n^2),$$

where $Z_n = X_n + Y_n$, $T_n = X_n - Y_n$.

Now as $X_n \xrightarrow[\text{in p}]{} a$, $Y_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$, by (iii) we get

$$Z_n \xrightarrow[\text{in p}]{} a + b, \quad T_n \xrightarrow[\text{in p}]{} a - b \text{ as } n \to \infty.$$

Therefore, $Z_n^2 \xrightarrow[\text{in p}]{} (a+b)^2$, $T_n^2 \xrightarrow[\text{in p}]{} (a-b)^2$ as $n \to \infty$ by (iv).

Therefore, $Z_n^2 - T_n^2 \xrightarrow[\text{in p}]{} (a+b)^2 - (a-b)^2 = 4ab$ as $n \to \infty$ by (iii).

Therefore, $\frac{1}{4}(Z_n^2 - T_n^2) \xrightarrow[\text{in p}]{} ab$ as $n \to \infty$ by (ii).

But $\frac{1}{4}(Z_n^2 - T_n^2) = X_n Y_n$.

Hence, $X_n Y_n \xrightarrow[\text{in p}]{} ab$, as $n \to \infty$.

(vi) For any given $\varepsilon > 0$, let A_n, B_n represent respectively the events

$$`|Y_n - b| < |b|', \quad `\left|\frac{1}{Y_n} - \frac{1}{b}\right| \geq \varepsilon'.$$

Now,
$$\left|\frac{1}{Y_n} - \frac{1}{b}\right| = \left|\frac{Y_n - b}{bY_n}\right| = \frac{|Y_n - b|}{|b||b + (Y_n - b)|}$$

$$\leq \frac{|Y_n - b|}{|b|\,\big||b| - |Y_n - b|\big|}, \tag{10.4.7}$$

since $|b + (Y_n - b)| \geq \big||b| - |Y_n - b|\big|$

If A_n and B_n occur simultaneously, then

$\big||b| - |Y_n - b|\big| = |b| - |Y_n - b|$, since $|Y_n - b| < |b|$,

and so by (10.4.7),

$$\frac{|Y_n - b|}{|b|(|b| - |Y_n - b|)} \geq \left|\frac{1}{Y_n} - \frac{1}{b}\right| \geq \varepsilon$$

or, $|Y_n - b| \geq \varepsilon |b|^2 - \varepsilon |b||Y_n - b|$, since $|b| - |Y_n - b| > 0$

or, $|Y_n - b|(1 + \varepsilon |b|) \geq \varepsilon |b|^2$

or, $|Y_n - b| \geq \dfrac{\varepsilon |b|^2}{1 + \varepsilon |b|} = \varepsilon'$ (say).

So if C_n be the event '$|Y_n - b| \geq \varepsilon'$', we get

$$A_n B_n \subseteq C_n.$$

Therefore, $B_n = A_n B_n + \bar{A}_n B_n \subseteq C_n + \bar{A}_n$.

Therefore, $0 \leq P(B_n) \leq P(C_n + \bar{A}_n) \leq P(C_n) + P(\bar{A}_n)$. (10.4.8)

Now since $Y_n \xrightarrow[\text{in p}]{} b$ as $n \to \infty$,

$$P(C_n) = P(|Y_n - b| \geq \varepsilon') \to 0 \text{ as } n \to \infty$$

and $P(\bar{A}_n) = P(|Y_n - b| \geq |b|) \to 0$ as $n \to \infty$.

Then from (10.4.8),

$$P(B_n) = P\left(\left|\frac{1}{Y_n} - \frac{1}{b}\right| \geq \varepsilon\right) \to 0 \text{ as } n \to \infty \text{ for every } \varepsilon > 0,$$

and hence $\dfrac{1}{Y_n} \xrightarrow[\text{in p}]{} \dfrac{1}{b}$ as $n \to \infty$.

Finally, $\dfrac{X_n}{Y_n} \xrightarrow[\text{in p}]{} \dfrac{a}{b}$ as $n \to \infty$ by (v).

Theorem 10.4.2. *If $X_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$ and $g : R \to R$ is a continuous function, then $g(X_n) \xrightarrow[\text{in p}]{} g(a)$ as $n \to \infty$.*

Proof : Since g is a continuous function for all $x \in R$, it is continuous at $x = a$ and so for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|x - a| < \delta$ implies $|g(x) - g(a)| < \varepsilon$.

Hence, $|X_n(\omega) - a| < \delta$ implies $|g\{X_n(\omega)\} - g(a)| < \varepsilon$, that is, the event '$|X_n - a| < \delta$' implies the event '$|g(X_n) - g(a)| < \varepsilon$'.

Therefore, $P(|X_n - a| < \delta) \le P\{|g(X_n) - g(a)| < \varepsilon\}$

or, $1 - P(|X_n - a| \ge \delta) \le 1 - P\{|g(X_n) - g(a)| \ge \varepsilon\}$

Therefore, $0 \le P\{|g(X_n) - g(a)| \ge \varepsilon\} \le P(|X_n - a| \ge \delta)$. (10.4.9)

Now since $X_n \xrightarrow[\text{in p}]{} a$ as $n \to \infty$,

$$\lim_{n \to \infty} P(|X_n - a| \ge \delta) = 0$$

Hence, proceeding to the limit as $n \to \infty$, we get from (10.4.9)

$$\lim_{n \to \infty} P\{|g(X_n) - g(a)| \ge \varepsilon\} = 0,$$

or, $g(X_n) \xrightarrow[\text{in p}]{} g(a)$ as $n \to \infty$.

Theorem 10.4.3. *If $X_n \xrightarrow[\text{in p}]{} X$ as $n \to \infty$, then $X_n \xrightarrow{d} X$ as $n \to \infty$, that is, convergence in probability implies convergence in distribution.*

Proof : Let $F_n(x)$ and $F(x)$ be the distribution functions of X_n and X respectively. Now for two real numbers a' and a with $a' < a$, we have

$$(X \le a') = (X_n \le a, X \le a') + (X_n > a, X \le a').$$

Again $(X_n \le a, X \le a') \subseteq (X_n \le a)$. So we get

$$(X \le a') \subseteq (X_n \le a) + (X_n > a, X \le a').$$

Therefore, $P(X \le a') \le P\{(X_n \le a) + (X_n > a, X \le a')\}$

$$\le P(X_n \le a) + P(X_n > a, X \le a')$$

or, $F(a') \le F_n(a) + P(X_n > a, X \le a')$. ... (10.4.10)

Now if $X_n > a$, $X \le a'$ occur simultaneously, then $X_n > a, -X \ge -a'$ and so $X_n - X > a - a'$.

So $\omega \in (X_n > a, X \le a') \Rightarrow \omega \in (X_n - X > a - a')$.

Again $\omega \in (X_n - X > a - a') \Rightarrow (|X_n - X| > a - a')$,

since $|(X_n - X)\omega| \ge (X_n - X)\omega$.

Therefore, $(X_n > a, X \le a') \subseteq (|X_n - X| > a - a')$.

So $P(X_n > a, X \le a') \le P(|X_n - X| > a - a')$.

Therefore, $0 \le P(X_n > a, X \le a') \le P(|X_n - X| > a - a')$. (10.4.11)

Now since $a - a' > 0$ and $X_n \xrightarrow[\text{in p}]{} X$ as $n \to \infty$, we get

$$\lim_{n \to \infty} P(|X_n - X| > a - a') = 0.$$

Hence from (10.4.11) we get

$$\lim_{n \to \infty} P(X_n > a, X \le a') = 0.$$

Now from (10.4.10) we have

$$\underline{\lim} F_n(a) \ge F(a') - \underline{\lim} P(X_n > a, X \le a').$$

But $\quad \underline{\lim} P(X_n > a, X \le a') = \overline{\lim} P(X_n > a, X \le a')$

$$= \lim_{n \to \infty} P(X_n > a, X \le a') = 0.$$

So we get $\underline{\lim} F_n(a) \ge F(a')$, where $a' < a$.

Similarly, considering the events $(X_n \le a)$, $(X \le a'')$, where $a'' > a$ and noting that

$$(X_n \le a) = (X \le a'', X_n \le a) + (X > a'', X_n \le a)$$

we find that $F_n(a) \le F(a'') + P(X > a'', X_n \le a)$,

where $\lim_{n \to \infty} P(X > a'', X_n \le a) = 0$

and consequently we get $\overline{\lim} F_n(a) \le F(a'')$.

Thus, for $a' < a < a''$ we have

$$F(a') \le \underline{\lim} F_n(a) \le \overline{\lim} F_n(a) \le F(a''). \quad \ldots \quad (10.4.12)$$

We know that the set of points of discontinuity of F is at most enumerable.

Now let F be continuous at a.

Then $\lim_{a' \to a-0} F(a') = \lim_{a'' \to a+0} F(a'') = F(a)$. So from (10.4.12) we get

$$\lim_{a' \to a-0} F(a') \le \underline{\lim}\, F_n(a) \le \overline{\lim}\, F_n(a) \le \lim_{a'' \to a+0} F(a'').$$

Therefore $F(a) \le \underline{\lim}\, F_n(a) \le \overline{\lim}\, F_n(a) \le F(a)$.

Hence we get $\quad \underline{\lim}\, F_n(a) = \overline{\lim}\, F_n(a) = F(a)$.

So $\lim_{n \to \infty} F_n(a)$ exists and is equal to $F(a)$.

Now a can be chosen arbitrarily such that F is continuous at a. Hence we can write $\lim_{n \to \infty} F_n(x) = F(x)$ at every point of continuity of F.

So $\quad X_n \xrightarrow{d} X$ as $n \to \infty$.

Note. The converse of Theorem 10.4.3 is not true in general, that is $X_n \xrightarrow{d} X$ does not imply $X_n \xrightarrow[\text{in p}]{} X$ as $n \to \infty$ in general. For, let us consider random variables $X, X_1, X_2, \ldots$ having identical distribution where the joint distribution of (X, X_n) is given by

$X \backslash X_n$	0	1	
0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	0	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	1

If $F_n(x)$ and $F(x)$ be the distribution functions of X_n and X repectively, then

$$\begin{aligned} F_n(x) = F(x) &= 0, \quad x < 0 \\ &= \frac{1}{2}, \quad 0 \le x < 1 \\ &= 1, \quad x \ge 1. \end{aligned}$$

Therefore, $\underset{n \to \infty}{Lt} F_n(x) = F(x)\ \forall\ x$, that is, $X_n \xrightarrow{d} X$ as $n \to \infty$.

But $\quad P\left(|X_n - X| > \frac{1}{2}\right) \ge P\left(|X_n - X| = 1\right)$, [since the event $(|X_n - X| = 1)$ implies the event $(|X_n - X| > \frac{1}{2})$]

$$= P(X_n = 0, X = 1) + P(X_n = 1, X = 0)$$

$$= \frac{1}{2} + \frac{1}{2} = 1.$$

Therefore, $\underset{n \to \infty}{Lt}\ P\left(|X_n - X| > \frac{1}{2}\right) \ne 0.$

Hence X_n does not tend to X in probability as $n \to \infty$.

Thus $X_n \xrightarrow{d} X$ does not imply $X_n \xrightarrow[\text{in p}]{} X$ as $n \to \infty$ in general.

10.5. Tchebycheff's Theorem, Bernoulli's Theorem, Law of Large Numbers.

Theorem 10.5.1. Tchebycheff's Theorem : *If $\{X_n\}$ be a sequence of random variables such that for any n, X_n has a finite mean m_n and finite standard deviation σ_n, then*

$$X_n - m_n \xrightarrow[\text{in p}]{} 0 \text{ as } n \to \infty,$$

provided $\quad \lim_{n \to \infty} \sigma_n = 0.$

Proof: Since $\quad E(X_n - m_n) = 0$ and

$$\text{var}(X_n - m_n) = \text{var}\, X_n = \sigma_n^2,$$

we have, by Tchebycheff's inequality,

$$0 \le P(|X_n - m_n| \ge \varepsilon) \le \frac{\sigma_n^2}{\varepsilon^2} \text{ for any } \varepsilon > 0.$$

Now since $\sigma_n \to 0$ as $n \to \infty$, we get

$$\lim_{n \to \infty} P(|X_n - m_n| \ge \varepsilon) = 0 \text{ for every } \varepsilon > 0,$$

which implies that $X_n - m_n \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Note. If, in addition to the condition of the Theorem 10.5.1, we have $\lim_{n \to \infty} m_n = m$, then from Theorem 10.4.1 (*iii*) and Theorem 10.3.1, we get $X_n \xrightarrow[\text{in p}]{} m$ as $n \to \infty$.

Theorem 10.5.2. Bernoulli's Theorem.

If X_n is a binomial (n, p) variate., then $\dfrac{X_n}{n} \xrightarrow[\text{in p}]{} p$ *as* $n \to \infty$.

Proof: Let $Y_n = \dfrac{X_n}{n}$.

Since X_n is a binomial (n, p) variate,

$$E(X_n) = np \text{ and } \operatorname{var} X_n = np(1-p).$$

Then
$$E(Y_n) = \frac{1}{n} E(X_n) = \frac{1}{n} \cdot np = p,$$

and
$$\operatorname{var} Y_n = \operatorname{var}\left(\frac{X_n}{n}\right) = \frac{1}{n^2} \operatorname{var} X_n = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Let $\varepsilon > 0$ be any given number. Then by Tchebycheff's inequality,

$$P(|Y_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Therefore, $0 \leq P(|Y_n - p| \geq \varepsilon) \leq \dfrac{p(1-p)}{n\varepsilon^2}$.

Now
$$\operatorname*{Lt}_{n \to \infty} \frac{p(1-p)}{n\varepsilon^2} = 0.$$

Therefore,
$$\operatorname*{Lt}_{n \to \infty} P(|Y_n - p| \geq \varepsilon) = 0,$$

or,
$$\operatorname*{Lt}_{n \to \infty} P\left(\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right) = 0.$$

Therefore, $\dfrac{X_n}{n} \xrightarrow[\text{in p}]{} p$ as $n \to \infty$.

Note. Bernoulli's theorem provides a logical basis of the frequency interpretation of probability. Let E be a given random experiment and let A be an event connected to E. Let E be repeated n times under identical conditions. Then we get n independent trials of E and these trials form Bernoullian sequence of n trials if we consider the event 'A' as 'success' and the event 'not A' as failure. If X_n be the random variable denoting the number of successes in this case, then the frequency ratio $f(A)$ of the event A is $\dfrac{X_n}{n}$, where X_n is a binomial $(n, p(A))$ variate. Now by Bernoulli's theorem

$$\frac{X_n}{n} \xrightarrow[\text{in p}]{} P(A) \text{ as } n \to \infty.$$

Then from the definition of convergence in probability we can state that the event "values of $\dfrac{X_n}{n}$ lie in a small neighbourhood of $P(A)$" is nearly certain if n is very large. So if n is very large, the event '$\dfrac{X_n}{n} \simeq P(A)$' is almost certain that is, '$f(A) \simeq P(A)$' is nearly certain if n is very large. So, by Bernoulli's theorem, we get a logical basis of the frequency interpretation of probability, which states that $f(A) \simeq P(A)$ if n is very large.

Theorem 10.5.3. Law of Large Numbers.

Let $\{X_n\}$ be a sequence of random variables such that $S_n = X_1 + X_2 + \cdots + X_n$ has a finite mean M_n and a finite variance B_n for all n. Then $\dfrac{S_n - M_n}{n} \xrightarrow[\text{in p}]{} 0$ *as* $n \to \infty$ *if* $\dfrac{B_n}{n^2} \to 0$ *as* $n \to \infty$.

Proof: Let ε be any positive number.

Now
$$E\left(\frac{S_n - M_n}{n}\right) = \frac{1}{n}\{E(S_n) - M_n\} = 0$$

and
$$\operatorname{var}\left(\frac{S_n - M_n}{n}\right) = \frac{1}{n^2} \operatorname{var} S_n = \frac{B_n}{n^2}.$$

Now by Tchebycheff's inequality,

$$P\left(\left|\frac{S_n - M_n}{n} - 0\right| \geq \varepsilon\right) \leq \frac{B_n}{n^2 \varepsilon^2}. \tag{10.5.1}$$

Now by the given condition $\displaystyle\lim_{n \to \infty} \frac{B_n}{n^2} = 0$, and so

$$\lim_{n \to \infty} \frac{B_n}{n^2 \varepsilon^2} = \frac{1}{\varepsilon^2} \lim_{n \to \infty} \frac{B_n}{n^2} = 0.$$

Hence, proceeding to the limit as $n \to \infty$, we get, from (10.5.1),

$$\lim_{n \to \infty} P\left(\left|\frac{S_n - M_n}{n}\right| \geq \varepsilon\right) = 0 \text{ for any } \varepsilon > 0,$$

which proves that $\dfrac{S_n - M_n}{n} \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

If for any n, $E(X_n)=m_n$, $M_n=m_1+m_2+\cdots+m_n$, then the above theorem can be written in the alternate form:

$$\bar{X}-\bar{m} \xrightarrow[\text{in } p]{} 0 \text{ as } n\to\infty,$$

where
$$\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{S_n}{n}$$

and
$$\bar{m}=\frac{m_1+m_2+\cdots+m_n}{n}=\frac{M_n}{n}.$$

Note. The above Law of Large Numbers is also known as Weak Law of Large Numbers (W.L.L.N).

Theorem 10.5.4. Law of Large Numbers for Equal Components. *If the random variables $X_1, X_2, \ldots, X_n, \ldots$ have the same distribution with finite mean m and finite standard deviation σ and if $X_1, X_2, \ldots, X_n$ are mutually independent for all n, then*

$$\bar{X} \xrightarrow[\text{in } p]{} m \quad \text{as } n\to\infty$$

where
$$\bar{X}=\frac{X_1+X_2+\cdots+X_n}{n}.$$

Proof: We have
$$E(\bar{X})=\frac{1}{n}E(X_1+X_2+\cdots+X_n)$$
$$=\frac{1}{n}\{E(X_1)+E(X_2)+\cdots+E(X_n)\}$$
$$=\frac{1}{n}.nm=m,$$

and since $X_1, X_2, \ldots, X_n$ are mutually independent,

$$\text{var}\,\bar{X}=\frac{1}{n^2}(\text{var}\,X_1+\text{var}\,X_2+\cdots+\text{var}\,X_n)$$
$$=\frac{1}{n^2}.n\sigma^2=\frac{\sigma^2}{n}.$$

Let $\varepsilon>0$ be given. By Tchebycheff's inequality,

$$P(|\bar{X}-m|\geq\varepsilon)\leq\frac{\frac{\sigma^2}{n}}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}.$$

Therefore,
$$0\leq P(|\bar{X}-m|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}$$

and since $\lim_{n\to\infty}\frac{\sigma^2}{n\varepsilon^2}=0$, we get

$$\lim_{n\to\infty}P(|\bar{X}-m|\geq\varepsilon)=0 \quad \text{for every } \varepsilon>0.$$

So $\bar{X} \xrightarrow[\text{in } p]{} m$ as $n\to\infty$.

Note 1. We have proved the Law of Large Numbers for Equal Components on the assumption that the common variance of the random variables X_n is finite, but the assumption is not necessary. It can be shown that the law holds under the condition that mean m of the common distribution exists finitely.

2. Here we note that the condition $\lim_{n\to\infty}\frac{B_n}{n^2}=0$ in Theorem 10.5.3 is satisfied.

We now show that Bernoulli's Theorem can be obtained as a particular case of Law of Large Numbers for Equal Components. We consider a sequence of mutually independent random variables $\{X_n\}$, where each X_n is a binomial $(1, p)$ variate. Then by the reproductive property of binomial variates

$$Y_n=X_1+X_2+\cdots+X_n$$

is a binomial (n, p) variate. Then for all n,

$$E(X_n)=1.p=p \text{ and var}\,X_n=1.p(1-p)=p(1-p).$$

Then by the Law of Large Numbers for Equal Components (Theorem 10.5.4),

$$\frac{Y_n}{n}\xrightarrow[\text{in } p]{} p \text{ as } n\to\infty.$$

where Y_n is a binomial (n, p) variate and so Bernoulli's Theorem is deduced.

We conclude this chapter by explaining the concept of asymptotic distribution and three fundamental limit theorems, namely, the Central Limit Theorem, De Moivre- Laplace Limit Theorem and the limit theorem for characteristic functions.

10.6. Asymptotic Distribution, Limit Theorem for Characteristic Functions, Central Limit Theorem and De Moivre- Laplace Limit Theorem.

Asymptotic Distribution: Let $\{X_n\}$ be a sequence of random variables such that $\{X_n\}$ converges in distribution to the random variable X with distribution function F, that is, $\underset{n\to\infty}{Lt}F_n(x)=F(x)$ at every point of continuity x of F, where F_n is the distribution function of X_n for $n=1,2,3,\ldots$. We express the aforesaid convergence by saying that X_n has the asymptotic distribution determined by the distribution function F.

In particular, if X has Poisson distribution with parameter μ, we say that X_n is asymptotically a Poisson variate with parameter μ. We take another example of asymptotic distribution where we use a terminology explained below.

Let $\{X_n\}$ be a sequence of random variables and $\{t_n\}$, $\{s_n\}$ be two sequences of real constants such that $s_n \neq 0$ for all n. Also let $F_n(x)$ be the distribution function of $\dfrac{X_n - t_n}{s_n}$. If $\lim\limits_{n\to\infty} F_n(x) = \Phi(x)$ for all x [$\Phi(x)$ is the distribution function of a standard normal variate and we note that $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$ is continuous for all x], then we say that $\dfrac{X_n - t_n}{s_n}$ is asymptotically a standard normal variate and in this case, for convenience, we say that X_n is *asymptotically normal* (t_n, s_n). We note that t_n and s_n are respectively not necessarily the mean and the standard deviation of X_n and further we observe that the practical sense in the statement 'X_n is asymptotically normal (t_n, s_n)' is that the distribution of $\dfrac{X_n - t_n}{s_n}$ is approximately normal $(0, 1)$ for large values of n and consequently X_n is approximately normal (t_n, s_n) for large values of n.

We state below an important theorem regarding convergence in distribution (without proof) known as the *Limit Theorem of Characteristic Functions*.

Theorem 10.6.1. *Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of random variables having distribution functions $F_1(x), F_2(x), \ldots, F_n(x), \ldots$ and the characteristic functions $\phi_1(t), \phi_2(t), \ldots, \phi_n(t), \ldots$ respectively. A necessary and sufficient condition that $\mathop{Lt}\limits_{n\to\infty} F_n(x) = F(x)$ at every point of continuity x of F is that, for any real t, $\mathop{Lt}\limits_{n\to\infty} \phi_n(t) = \phi(t)$, where $\phi(t)$ is continuous at $t = 0$ and $\phi(t)$ is the characteristic function of the distribution determined by the distribution function F.*

An Application of the above Theorem.

Poisson Distribution as a Limit of Binomial Distribution:

Let X_n be the binomial (n, p) variate, where $0 < p < 1$ and n is a positive integer. If $\phi_n(t)$ be the characteristic function of X_n, then

$$\phi_n(t) = E(e^{itX_n}) = (pe^{it} + 1 - p)^n, \text{ where } i = \sqrt{-1}.$$

Now for every real t,

$$\mathop{Lt}\limits_{n\to\infty} \phi_n(t) = \mathop{Lt}\limits_{n\to\infty} (pe^{it} + 1 - p)^n$$

$$= \mathop{Lt}\limits_{n\to\infty} \left(\frac{\mu}{n} e^{it} + 1 - \frac{\mu}{n}\right)^n, \text{ where } \mu = np\,(>0),$$

μ being a fixed positive number

$$= \mathop{Lt}\limits_{n\to\infty} \left\{1 + \frac{\mu}{n}(e^{it} - 1)\right\}^n$$

$$= \mathop{Lt}\limits_{n\to\infty} \left[\left\{1 + \frac{\mu(e^{it} - 1)}{n}\right\}^{\frac{n}{\mu(e^{it}-1)}}\right]^{\mu(e^{it}-1)}$$

$$= e^{\mu(e^{it}-1)}, \quad \text{since } \mathop{Lt}\limits_{n\to\infty}\left(1 + \frac{x}{n}\right)^{\frac{n}{x}} = e, \text{ if } x \neq 0,$$

where the function determined by $\phi(t) = e^{\mu(e^{it}-1)}$ is the characteristic function of a Poisson variate. Also $e^{\mu(e^{it}-1)}$ is continuous at $t = 0$. Hence, by the limit theorem of characteristic functions, the distribution function of X_n will tend to that of Poisson-μ variate as $n \to \infty$, if $np = \mu$, a positive constant and so we can state that Poisson distribution can be obtained as the limit of binomial (n, p) distribution as $n \to \infty$, when np is kept fixed.

At the beginning of this section we introduced the concept of asymptotic distribution and in particular we noted that if the sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ be such that X_n is asymptotically normal (t_n, s_n), then we can say that X_n is approximately normal (t_n, s_n) for large values of n. It is surprising to note that the aforesaid behaviour is observed for a large class of sequences of random variables having certain properties and this is expressed by the theorem known as 'Central Limit Theorem'. We shall not state the general form of this theorem but a particular form known as 'Central Limit Theorem for Equal Components'.

Central Limit Theorem for Equal Components.

Theorem 10.6.2. *Let $\{X_n\}$ be a sequence of random variables, where $X_1, X_2, \ldots, X_n, \ldots$ all have the same distribution with common mean m and common standard deviation $\sigma\,(>0)$ and if $X_1, X_2, \ldots, X_n$ are mutually independent for all n, then $\dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$ is asymptotically normal*

(0,1) or $\overline{X}$ is asymptotically normal $\left(m, \frac{\sigma}{\sqrt{n}}\right)$, where $\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$, that is, $\underset{n\to\infty}{Lt} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ for all x,

$F_n(x)$ being the distribution function of $\frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$ for $n = 1, 2, \ldots$

Proof : Here $X_1, X_2, \ldots, X_n, \ldots$ have the same distribution with mean m and standard deviation $\sigma (>0)$ and so $\frac{X_1 - m}{\sigma}, \frac{X_2 - m}{\sigma}, \ldots, \frac{X_n - m}{\sigma}, \ldots$, have the same distribution and consequently each $\frac{X_r - m}{\sigma}$ $(r = 1, 2, \ldots, n, \ldots)$ has the same characteristic function, say $\phi(t)$.

Then $E\left\{e^{\frac{it(X_r - m)}{\sigma}}\right\} = \phi(t), \quad r = 1, 2, \ldots n, \ldots\ldots$ (10.6.1)

Let $$Y_n = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}.$$

Then $$Y_n = \frac{\frac{X_1 - m}{\sigma} + \frac{X_2 - m}{\sigma} + \cdots + \frac{X_n - m}{\sigma}}{\sqrt{n}},$$

where $\frac{X_1 - m}{\sigma}, \frac{X_2 - m}{\sigma}, \ldots, \frac{X_n - m}{\sigma}$ are mutually independent, since $X_1, X_2, \ldots, X_n$ are mutually independent.

So the characteristic function $\phi_n(t)$ of Y_n is given by

$$\phi_n(t) = E\left\{e^{\frac{it}{\sqrt{n}}\left(\frac{X_1 - m}{\sigma}\right)}\right\} E\left\{e^{\frac{it}{\sqrt{n}}\left(\frac{X_2 - m}{\sigma}\right)}\right\} \ldots E\left\{e^{\frac{it}{\sqrt{n}}\left(\frac{X_n - m}{\sigma}\right)}\right\}.$$

Then by (10.6.1),

$$\phi_n(t) = \left\{\phi\left(\frac{t}{\sqrt{n}}\right)\right\}^n. \quad (10.6.2)$$

Here $E\left(\frac{X_r - m}{\sigma}\right) = 0$ and $E\left(\frac{X_r - m}{\sigma}\right)^2 = 1$ for $r = 1, 2, \ldots, n, \ldots$

Now $\phi(t)$ is a complex valued function of the real variable t such that $\phi(t)$ has a finite second order derivative at $t = 0$ and the derivatives $\phi'(0)$, $\phi''(0)$ are given by

$$\phi'(0) = i\, E\left(\frac{X_r - m}{\sigma}\right) = 0$$

and $$\phi''(0) = i^2 E\left(\frac{X_r - m}{\sigma}\right)^2 = -1,$$

since $E(X_r - m) = 0$ and $E(X_r - m)^2 = \sigma^2$ for each r.

Now by a particular form of Taylor's Theorem [Hardy - Pure Mathematics (10-th edition) page 290], we have

$$\phi(t) = \phi(0) + t\,\phi'(0) + \frac{t^2}{2!}\phi''(0) + O(t^2), \text{ where } \underset{t\to 0}{Lt} \frac{O(t^2)}{t^2} = 0.$$

Then we get

$$\phi(t) = 1 - \frac{t^2}{2!} + O(t^2), \quad \text{since } \phi(0) = 1. \quad \ldots \quad (10.6.3)$$

So from (10.6.2) and (10.6.3) we get

$$\phi_n(t) = \left\{1 - \frac{t^2}{2n} + O\left(\frac{t^2}{n}\right)\right\}^n. \quad \ldots \quad (10.6.4)$$

At this stage, we now state below an important theorem on limit (which we shall use) on sequence of complex numbers.

If the sequence $\{C_n\}$ of complex numbers be convergent to the limit c, then

$$\underset{n\to\infty}{Lt}\left(1 + \frac{C_n}{n}\right)^n = e^c. \quad \ldots \quad (10.6.5)$$

Now $1 - \frac{t^2}{2n} + O\left(\frac{t^2}{n}\right)$ can be expressed as $1 + \frac{C_n}{n}$, where

$C_n = -\frac{t^2}{2} + n\,O\left(\frac{t^2}{n}\right)$ and $\underset{n\to\infty}{Lt} C_n = -\frac{t^2}{2}$, since

$$\underset{n\to\infty}{Lt}\, n\,O\left(\frac{t^2}{n}\right) = \underset{n\to\infty}{Lt}\left\{t^2 \frac{O\left(\frac{t^2}{n}\right)}{\frac{t^2}{n}}\right\}$$

$$= \underset{u\to 0}{Lt}\; t^2 \frac{O(u^2)}{u^2} \quad \left(u = \frac{t}{\sqrt{n}}\right)$$

$= 0$, for fixed t.

Then by (10.6.5) we have

$$\underset{n\to\infty}{Lt}\left\{1-\frac{t^2}{2n}+O\left(\frac{t^2}{n}\right)\right\}^n=e^{-\frac{t^2}{2}}$$

and hence,by (10.6.4) we get

$$\underset{n\to\infty}{Lt}\ \phi_n(t)=e^{-\frac{t^2}{2}},$$

which gives the characteristic function of a standard normal variate. So by the limit theorem of characteristic function we conclude that $Y_n=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ converges in distribution to a standard normal variate as $n\to\infty$, that is, $\overline{X}$ is asymptotically normal $\left(m,\dfrac{\sigma}{\sqrt{n}}\right)$.

Remark. Let $\{X_n\}$ be a sequence of mutually independent random variables such that each X_n has an identical Cauchy distribution with parameters (1, 0). Then the characteristic function of each X_n is $e^{-|t|}$ and so that of $\overline{X}=\dfrac{S_n}{n}$ is also $e^{-|t|}$. It is then evident that Central Limit Theorem does not hold for this sequence. This is because of the fact tha. Cauchy distribution does not possess finite mean and variance.

Central Limit Theorem (for Equal Components) implies Law of Large Numbers (for Equal Components).

Let $\{X_n\}$ be a sequence of mutually independent random variables such that each X_n has the same distribution with finite mean m and finite standard deviation σ.

Let $\quad Y_n=\dfrac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}$ where $\overline{X}=\dfrac{X_1+X_2+\cdots+X_n}{n}$.

If $F_n(x)$ be the distribution function of Y_n, then by Central Limit Theorem, $\underset{n\to\infty}{Lt}\ F_n(x)=\Phi(x)$ for all x where $\Phi(x)$ is the distribution function of a standard normal variate.

Let b be any given positive number. Then for any $\varepsilon>0$,

$$P(|\overline{X}-m|\ge\varepsilon)=P\left(\left|\frac{\overline{X}-m}{\frac{\sigma}{\sqrt{n}}}\right|\ge\sqrt{n}\frac{\varepsilon}{\sigma}\right)=P\left(|Y_n|\ge\frac{\sqrt{n}\,\varepsilon}{\sigma}\right).$$

Now $\quad |Y_n|\ge\dfrac{\sqrt{n}\,\varepsilon}{\sigma}\Rightarrow|Y_n|>b$ if $\dfrac{\sqrt{n}\,\varepsilon}{\sigma}>b$.

Therefore $\quad P\left(|Y_n|\ge\dfrac{\sqrt{n}\,\varepsilon}{\sigma}\right)\le P\left(|Y_n|>b\right)$

if $\dfrac{\sqrt{n}\,\varepsilon}{\sigma}>b$, that is, if $n>\dfrac{b^2\sigma^2}{\varepsilon^2}$.

Now $\quad P(|Y_n|>b)=1-P(|Y_n|\le b)$

$$=1-P(-b\le Y_n\le b)\le1-P(-b<Y_n\le b),$$

since $(-b<Y_n\le b)$ is a subevent of $(-b\le Y_n\le b)$.

Therefore $P(|Y_n|>b)\le1-\{F_n(b)-F_n(-b)\}$.

Hence we get

$$P\left(|Y_n|\ge\frac{\sqrt{n}\,\varepsilon}{\sigma}\right)\le1-F_n(b)+F_n(-b)\ \text{ if } n>\frac{b^2\sigma^2}{\varepsilon^2}.$$

So $\quad P(|\overline{X}-m|\ge\varepsilon)\le1-F_n(b)+F_n(-b)$ if $n>\dfrac{b^2\sigma^2}{\varepsilon^2}$. (10.6.6)

Hence proceeding to the limit as $n\to\infty$, we get, from (10.6.6),

$\underset{n\to\infty}{Lt}\ P(|\overline{X}-m|\ge\varepsilon)\le\Phi(-b)+\{1-\Phi(b)\}$, that is,

$\underset{n\to\infty}{Lt}\ P(|\overline{X}-m|\ge\varepsilon)\le2\{1-\Phi(b)\}$, since $\Phi(-b)=1-\Phi(b)$.

Therfore $0\le\underset{n\to\infty}{Lt}\ P(|\overline{X}-m|\ge\varepsilon)\le2\{1-\Phi(b)\}$. (10.6.7)

This is true for any $b(>0)$ and so again proceeding to the limit as $b\to\infty$, we get, from (10.6.7),

$$\underset{n\to\infty}{Lt}\ P(|\overline{X}-m|\ge\varepsilon)=0,\text{ since }\underset{b\to\infty}{Lt}\ \{1-\Phi(b)\}=0.$$

Hence $\overline{X}\xrightarrow[\text{in p}]{}m$ as $n\to\infty$ and hence the Law of Large Numbers for Equal Components is deduced.

De Moivre - Laplace Limit Theorem.

Theorem 10.6.3. *If X_n be a binomial (n, p) variate for every n (p is a given number such that $0<p<1$), then the sequence $\left\{\dfrac{X_n-np}{\sqrt{np(1-p)}}\right\}$ is convergent in distribution to a standard normal variate as $n\to\infty$, that is,*

$$\underset{n\to\infty}{Lt}\ F_n(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}\,dt$$

for all real values of x, where $F_n(x)$ is the distribution function of $\dfrac{X_n-np}{\sqrt{np(1-p)}}$ for $n=1,2,3,\ldots$.

Proof: By the reproductive property of binomial distribution we can say that for each positive integer n, the binomial (n, p) variate X_n can be expressed as

$$X_n = Y_1 + Y_2 + \cdots + Y_n,$$

where $Y_1, Y_2, \ldots, Y_n$ are mutually independent binomial $(1, p)$ variates. Then $Y_1, Y_2, \ldots, Y_n, \ldots$ have the same distribution with mean p and standard deviation $\sqrt{p(1-p)}$ (>0) and further $Y_1, Y_2, \ldots, Y_n$ are mutually independent for all n. So by the Central Limit Theorem for Equal Components we find that

$$\frac{\frac{Y_1 + Y_2 + \cdots + Y_n}{n} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}}$$

is asymptotically normal (0, 1).

Now $$\frac{\frac{Y_1 + Y_2 + \cdots + Y_n}{n} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} = \frac{X_n - np}{\sqrt{np(1-p)}}.$$

So $\left\{\frac{X_n - np}{\sqrt{np(1-p)}}\right\}$ is convergent in distribution to a standard normal variate as $n \to \infty$, that is, $\underset{n\to\infty}{Lt} F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ for all real values of x, $F_n(x)$ being the distribution function of $\frac{X_n - np}{\sqrt{np(1-p)}}$.

Normal Approximation to Binomial Distribution.

From De Moivre – Laplace Limit Theorem we observe that if n be sufficiently large and p is fixed ($0 < p < 1$), then the distribution of the binomial (n, p) variate X_n is approximately normal $(np, \sqrt{np(1-p)})$, since that of $\frac{X_n - np}{\sqrt{np(1-p)}}$ is approximately normal $(0, 1)$.

10.7. Illustrative Examples.

Ex. 1. *The distribution of a random variable X is given by*

$$P(X = -1) = \frac{1}{8},\ P(X = 0) = \frac{3}{4},\ P(X = 1) = \frac{1}{8}.$$

Verify Tchebycheff's inequality for the distribution.

[C. H. (Math.) '79, 84]

Here mean m of the given distribution is

$$m = E(X) = -1.\frac{1}{8} + 0.\frac{3}{4} + 1.\frac{1}{8} = 0$$

and variance σ^2 is given by

$$\sigma^2 = \operatorname{var} X = E(X^2) - m^2$$
$$= (-1)^2\frac{1}{8} + 0^2.\frac{3}{4} + 1^2.\frac{1}{8} = \frac{1}{4}.$$

We are to show that for any $\varepsilon > 0$,

$$P(|X - 0| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2},$$

that is, $$P(|X| \geq \varepsilon) \leq \frac{1}{4\varepsilon^2}.$$

Now we have the following possibilities :

(*i*) $0 < \varepsilon \leq 1$, (*ii*) $\varepsilon > 1$.

Case (i). In this case $P(|X| \geq \varepsilon) = P(X = -1) + P(X = 1)$

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Again $\frac{1}{4\varepsilon^2} \geq \frac{1}{4}$, since here $0 < \varepsilon \leq 1$.

Therefore, $P(|X| \geq \varepsilon) = \frac{1}{4} \leq \frac{1}{4\varepsilon^2}$.

Case (ii). Here $\varepsilon > 1$.

Therefore, $P(|X| \geq \varepsilon) = 0 < \frac{1}{4\varepsilon^2}$ since here $(|X| \geq \varepsilon)$ is an impossible event.

Hence Tchebycheff's inequality is verified.

Ex. 2. *Show by Tchebycheff's inequality that in 2000 throws with a coin the probability that the number of heads lies between 900 and 1100 is at least* $\frac{19}{20}$.

[C. H. (Math.) '69]

Let X be the random variable denoting the number of heads in 2000 throws of a coin. Then X is a binomial $\left(2000, \frac{1}{2}\right)$ variate.

So $$E(X) = 2000 \times \frac{1}{2} = 1000$$

and $$\operatorname{var} X = 2000 \times \frac{1}{2} \times \frac{1}{2} = 500.$$

Now $P(900 < X < 1100) = P(-100 < X - 1000 < 100)$
$= P(|X - 1000| < 100)$
$= 1 - P(|X - 1000| \geq 100)$.

Then by Tchebycheff's inequality, taking $\varepsilon = 100$,

$$P(|X - 1000| \geq 100) \leq \frac{\text{var } X}{(100)^2} = \frac{500}{(100)^2} = \frac{1}{20}.$$

Therefore, $1 - P(|X - 1000| \geq 100) \geq 1 - \frac{1}{20} = \frac{19}{20}$.

Therefore, $P(900 < X < 1100) \geq \frac{19}{20}$.

Ex. 3. *If $\{X_i\}_i$ be a sequence of independent random variables such that for each i, $E(X_i) = m_i$, var $X_i = \sigma_i^2 \leq \sigma^2 < \infty$; use Tchebycheff's inequality to show that*

$$\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n} \xrightarrow[\text{in p}]{} 0 \quad \text{as } n \to \infty. \qquad [\text{C. H. (Math.) '80, '86}]$$

Here $E(X_i) = m_i$ and var $X_i = \sigma_i^2$, $\quad i = 1, 2, \ldots, n, \ldots$

Now $E\left(\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n}\right)$

$$= \frac{m_1 + m_2 + \cdots + m_n}{n} - \frac{m_1 + m_2 + \cdots + m_n}{n}$$

$= 0$.

$\text{var}\left(\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n}\right)$

$= \text{var}\left\{\frac{1}{n} \sum_{i=1}^{n} (X_i - m_i)\right\}$

$= \frac{1}{n^2}\{\text{var}(X_1 - m_1) + \text{var}(X_2 - m_2) + \cdots + \text{var}(X_n - m_n)\}$,

since $X_1, X_2, \ldots, X_n$ are mutually independent

$= \frac{1}{n^2} \sum_{i=1}^{n} \text{var } X_i$.

Now for each positive number ε, by Tchebycheff's inequality,

$$P\left\{\left|\left(\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n}\right) - 0\right| \geq \varepsilon\right\}$$

$$\leq \frac{\sum_{i=1}^{n} \text{var } X_i}{n^2 \varepsilon^2} = \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2 \varepsilon^2}$$

$$\leq \frac{n\sigma^2}{n^2 \varepsilon^2} = \frac{\sigma^2}{n \varepsilon^2}.$$

Also $\quad P\left\{\left|\left(\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n}\right)\right| \geq \varepsilon\right\} \geq 0.$

Therefore, $0 \leq P\left\{\left|\left(\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n}\right) - 0\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$.

Now $\underset{n \to \infty}{Lt} \frac{\sigma^2}{n\varepsilon^2} = 0$, since σ is finite and $\varepsilon > 0$ is given.

Therefore, $\underset{n \to \infty}{Lt} P\left\{\left|\left(\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n}\right) - 0\right| \geq \varepsilon\right\} = 0.$

Hence, $\sum_{i=1}^{n} \frac{X_i}{n} - \sum_{i=1}^{n} \frac{m_i}{n} \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Ex. 4. *Show that the probability that the number of heads in 2000 throws with a fair coin lies between 900 and 1100 is $\doteq 2F(2\sqrt{5}) - 1$, where $F(x)$ denotes the standard normal distribution function. (Compare it with the lower limit given by Tchebycheff's inequality).* [C. H. (Math.) '81]

Let X be the random variable denoting the number of 'heads'. Then X is a binomial $\left(2000, \frac{1}{2}\right)$ variate. Then by De Moivre-Laplace Limit Theorem,

$$X^* = \frac{X - 2000 \times \frac{1}{2}}{\sqrt{2000 \times \frac{1}{2} \times \frac{1}{2}}} = \frac{X - 1000}{\sqrt{500}}$$

is approximately a standard normal variate.

Now $P(900 < X < 1100)$

$$= P(-100 < X - 1000 < 100)$$

$$= P\left(\frac{-100}{\sqrt{500}} < \frac{X-1000}{\sqrt{500}} < \frac{100}{\sqrt{500}}\right)$$

$$= P(-2\sqrt{5} < X^* < 2\sqrt{5})$$

$\simeq P(-2\sqrt{5} < U < 2\sqrt{5})$, where U is the standard normal variate.

Now $P(-2\sqrt{5} < U < 2\sqrt{5}) = F(2\sqrt{5}) - F(-2\sqrt{5})$

$$= F(2\sqrt{5}) - \{1 - F(2\sqrt{5})\}$$

$$= 2F(2\sqrt{5}) - 1,$$

where $$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

Therefore, $P(900 < X < 1100) \simeq 2F(2\sqrt{5}) - 1 = 0.99992$, using standard normal table.

Also, using Tchebycheff's inequality (see Ex.2),

$$P(900 < X < 1100) \geq \frac{19}{20} = 0.95,$$

and we see that $0.99992 > 0.95$, which supports Tchebycheff's inequality regarding the lower limit.

Ex. 5. *By applying the Central Limit Theorem to a sequence of random variables with Poisson distribution, prove that*

$$\lim_{n\to\infty} e^{-n} \sum_{r=0}^{n} \frac{n^r}{r!} = \frac{1}{2}. \qquad \text{[C. H. (Math.) '91]}$$

We consider a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ each having Poisson-1 distribution where $X_1, X_2, \ldots, X_n$ are mutually independent for all n. Then

$$\left.\begin{array}{l} \text{mean of } X_i = m = 1 \\ \text{and standard deviation of } X_i = \sigma = 1 \end{array}\right\} \text{ for } i = 1, 2, \ldots n, \ldots$$

(Mean $= \mu$ and standard deviation $= \sqrt{\mu}$ for a Poisson - μ variate.)

Hence, by Central Limit Theorem for Equal Components

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is asymptotically normal $\left(1, \frac{1}{\sqrt{n}}\right)$, that is, $\dfrac{\overline{X}-1}{\frac{1}{\sqrt{n}}}$ converges in distribution to U as $n \to \infty$, where U is a standard normal variate.

Therefore $$\mathop{Lt}_{n\to\infty} P\left(\frac{\overline{X}-1}{\frac{1}{\sqrt{n}}} \leq 0\right) = P(U \leq 0) = \frac{1}{2}. \qquad (10.7.1)$$

But $$P\left(\frac{\overline{X}-1}{\frac{1}{\sqrt{n}}} \leq 0\right) = P(\overline{X} \leq 1)$$

$$= P(X_1 + X_2 + \cdots + X_n \leq n). \qquad (10.7.2)$$

Now by reproductive property of Poisson distribution $X_1 + X_2 + \cdots + X_n$ is a Poisson-n variate. Hence,

$$P(X_1 + X_2 + \cdots + X_n \leq n) = \sum_{r=0}^{n} \frac{e^{-n} n^r}{r!}. \qquad (10.7.3)$$

Therefore, from (10.7.1), (10.7.2) and (10.7.3), we get

$$\mathop{Lt}_{n\to\infty} \sum_{r=0}^{n} \frac{e^{-n} n^r}{r!} = \frac{1}{2},$$

that is, $$\mathop{Lt}_{n\to\infty} e^{-n} \sum_{r=0}^{n} \frac{n^r}{r!} = \frac{1}{2}.$$

Ex. 6. *Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent random variables, where each X_i takes values $-1, 0, 1$ for $i = 2, 3, \ldots$ Given that $P(X_i = 1) = \frac{1}{i} = P(X_i = -1)$ and $P(X_i = 0) = 1 - \frac{2}{i}$ for $i > 1$ and X_1 takes the value 0 with $P(X_1 = 0) = 1$, examine if the Law of Large Numbers holds for this sequence.*

We have $E(X_i) = 0\left(1 - \frac{2}{i}\right) + 1 \cdot \frac{1}{i} + (-1) \cdot \frac{1}{i} = 0$ for $i > 1$

and $E(X_1) = 0.1 = 0$.

$$\text{Var } X_i = E(X_i^2) - 0 = 0\left(1 - \frac{2}{i}\right) + 1 \cdot \frac{1}{i} + 1 \cdot \frac{1}{i} = \frac{2}{i} \quad \text{for } i > 1$$

and $$\text{var } X_1 = E(X_1^2) - \{E(X_1)\}^2 = 0.$$

Let $S_n = X_1 + X_2 + \cdots + X_n$.

Then $M_n = E(S_n) = E(X_1 + X_2 + \cdots + X_n)$

$= E(X_1) + E(X_2) + \cdots + E(X_n)$

$= 0$

and $\Sigma_n^2 = \text{var } S_n = \sum_{i=1}^{n} \text{var } X_i = 2 \sum_{i=2}^{n} \frac{1}{i}$

$$= 2\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \text{ if } n > 1.$$

Hence, S_n has a finite mean 0 and finite variance $2\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)$ for all $n > 1$ and variance is 0 if $n = 1$.

$$\text{Now } \frac{\text{var } S_n}{n^2} = \frac{2\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)}{n^2} \to 0 \text{ as } n \to \infty.$$

[If $\{x_n\}$ has a limit l, then the sequence $\{y_n\}$, defined by $y_n = \frac{x_1 + x_2 + \cdots + x_n}{n}$, has the same limit l.

Take $x_n = \frac{1}{n}$, then $\underset{n \to \infty}{Lt}\ x_n = 0$. Hence by the above result

$$\frac{1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}}{n} \to 0 \text{ as } n \to \infty,$$

and so $\underset{n \to \infty}{Lt} \dfrac{\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}}{n} = 0$, since $\underset{n \to \infty}{Lt} \dfrac{1}{n} = 0$.]

Hence the Law of Large Numbers holds for the given sequence.

Ex. 7. *A random variable X has a density function f(x) given by*

$$f(x) = e^{-x}, \quad x \geq 0$$
$$= 0, \quad \textit{elsewhere}.$$

Show that Tchebycheff's inequality gives

$$P(|X - 1| \geq 2) \leq \frac{1}{4}$$

and show that actual probability is e^{-3}.

We have $E(X) = \int_0^\infty x e^{-x}\, dx = \Gamma(2) = 1 = m,$

$$E(X^2) = \int_0^\infty x^2 e^{-x}\, dx = \Gamma(3) = 2.$$

Therefore, var $X = E(X^2) - m^2 = 2 - 1 = 1 = \sigma^2$.

Now by Tchebycheff's inequality, for any $\varepsilon > 0$,

$$P(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Taking $\varepsilon = 2$, we get

$$P(|X - 1| \geq 2) \leq \frac{1}{4}.$$

Now $P(|X - 1| \geq 2) = 1 - P(|X - 1| < 2)$

$= 1 - P(-2 < X - 1 < 2)$

$= 1 - P(-1 < X < 3)$

$= 1 - P(-1 \leq X \leq 3)$, since X is a continuous variate

$= 1 - \int_{-1}^{3} f(x)\, dx$

$= 1 - \int_{-1}^{0} f(x)\, dx - \int_{0}^{3} f(x)\, dx$

$= 1 - 0 - \int_0^3 e^{-x}\, dx$

$= 1 - (1 - e^{-3})$

$= e^{-3}$

which is the actual probability.

Ex. 8. *Examine if the condition in the Law of Large Numbers is satisfied by the sequence of independent random variables $\{X_n\}$, where*

$$X_n = \pm\sqrt{2n - 1} \quad \textit{with probability } \frac{1}{2}.$$

We have $E(X_n) = \sqrt{2n-1} \cdot \frac{1}{2} - \sqrt{2n-1} \cdot \frac{1}{2} = 0$ for all n,

and var $X_n = E(X_n^2) - 0 = (2n - 1) \cdot \frac{1}{2} + (2n - 1) \cdot \frac{1}{2} = 2n - 1$ for all n.

Let $S_n = X_1 + X_2 + \cdots + X_n$.

Then $M_n = E(S_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = 0$,

$$\text{var } S_n = \sum_{i=1}^{n} \text{var } X_i = \sum_{i=1}^{n} (2i-1) = n^2.$$

Hence S_n has a finite mean 0 and finite variance n^2 for all n.

Therefore $\dfrac{\text{var } S_n}{n^2} = 1$ and $\underset{n\to\infty}{Lt} \dfrac{\text{var } S_n}{n^2} = 1 \neq 0$.

Hence the condition in the Law of Large Numbers does not hold good for the given sequence.

Ex. 9. *A random variable X has probability density function $12x^2(1-x)$, $(0<x<1)$. Compute $P(|X-m| \geq 2\sigma)$ and compare it with the limit given by Tchebycheff's inequality.*

$$E(X) = \int_0^1 12x^3(1-x)\,dx = \frac{3}{5}, \qquad E(X^2) = \int_0^1 12x^4(1-x)\,dx = \frac{2}{5}.$$

$$\text{var } X = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}.$$

Therefore $P(|X-m| \geq 2\sigma) = P\left(\left|X - \dfrac{3}{5}\right| \geq \dfrac{2}{5}\right)$

$$= 1 - P\left(\frac{3}{5} - \frac{2}{5} < X < \frac{3}{5} + \frac{2}{5}\right)$$

$$= 1 - P\left(\frac{1}{5} < X < 1\right) = 1 - 12\int_{\frac{1}{5}}^{1} x^2(1-x)\,dx$$

$$= 1 - 12\left\{\left(\frac{1}{3} - \frac{1}{4}\right) - \left(\frac{1}{375} - \frac{1}{2500}\right)\right\}$$

$$= \frac{17}{625}$$

Now taking $\varepsilon = 2\sigma > 0$, we get, by Tchebycheff's inequality $P(|X-m| \geq 2\sigma) \leq \dfrac{1}{4}$ and we note that $\dfrac{17}{625} < \dfrac{1}{4}$, so that the result supports Tchebycheff's inequality regarding the upper limit.

Ex. 10. *Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables, each having mean m and variance σ^2. Then prove that $\sum_{i=1}^{n} \dfrac{X_i}{n} \to m$ in mean square as $n \to \infty$.* [C. H. (Math.) '91]

Here $E\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right) = \dfrac{mn}{n} = m$

and $\text{var}\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n}\right)$

$$= \frac{1}{n^2}(\text{var } X_1 + \text{var } X_2 + \cdots + \text{var } X_n)$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Then $E\left\{\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n} - m\right)^2\right\}$

$$= \text{var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

Now $\dfrac{\sigma^2}{n} \to 0$ as $n \to \infty$.

Therefore, $\underset{n\to\infty}{Lt}\; E\left\{\left(\dfrac{X_1 + X_2 + \cdots + X_n}{n} - m\right)^2\right\} = 0.$

So $\dfrac{X_1 + X_2 + \cdots + X_n}{n}$ is convergent in mean square to m as $n \to \infty$.

Ex. 11. *Let X_n be the random variable denoting the frequency ratio of successes in a Poisson sequence of n trials, where p_i is the probability of success in the i th trial $(i = 1, 2, \ldots, n)$. If $\bar{p} = \dfrac{p_1 + p_2 + \cdots + p_n}{n}$, then prove that*

$$X_n - \bar{p} \xrightarrow[\text{in p}]{} 0 \text{ as } n \to \infty.$$

We consider the sequence $\{Y_n\}$ of random variables, where

$Y_i = 1$ if we get success in the ith trial of the given Poisson trials

$= 0$ if we get failure in the ith trial of the given Poisson sequence for $i = 1, 2, 3, \ldots$.

Then $X_n = \dfrac{Y_1 + Y_2 + \cdots + Y_n}{n}$, where $Y_1, Y_2, \ldots, Y_n$ are mutually independent for all n.

Here the sepctrum of Y_i is $\{0, 1\}$, where

$P(Y_i = 0) = 1 - p_i$, $P(Y_i = 1) = p_i$ for $i = 1, 2, \ldots$

So $E(Y_i) = 0(1-p_i) + 1 \cdot p_i = p_i$

and $\operatorname{var} Y_i = E(Y_i^2) - \{E(Y_i)\}^2$

$= 0^2 \cdot (1-p_i) + 1^2 \cdot p_i - p_i^2$

$= p_i(1-p_i)$ for $i = 1, 2, \ldots$

So $E(Y_1 + Y_2 + \cdots + Y_n) = E(Y_1) + E(Y_2) + \cdots + E(Y_n)$

$= p_1 + p_2 + \cdots + p_n$

and $\operatorname{var}(Y_1 + Y_2 + \cdots + Y_n)$

$= \operatorname{var} Y_1 + \operatorname{var} Y_2 + \cdots + \operatorname{var} Y_n$

since $Y_1, Y_2, \ldots, Y_n$ are mutually independent

$= p_1(1-p_1) + p_2(1-p_2) + \cdots + p_n(1-p_n)$.

Now $\dfrac{p_i + 1 - p_i}{2} \geq \sqrt{p_i(1-p_i)}$, $(p_i > 0, 1 - p_i > 0)$.

Therefore, $p_i(1-p_i) \leq \frac{1}{4}$ for $i = 1, 2, 3, \ldots$

So $p_1(1-p_1) + p_2(1-p_2) + \cdots + p_n(1-p_n) \leq \frac{n}{4}$.

Then $\dfrac{\operatorname{var}(Y_1 + Y_2 + \cdots + Y_n)}{n^2} \leq \dfrac{1}{4n}$.

Also $\dfrac{\operatorname{var}(Y_1 + Y_2 + \cdots + Y_n)}{n^2} \geq 0$.

So we have $0 \leq \dfrac{\operatorname{var}(Y_1 + Y_2 + \cdots + Y_n)}{n^2} \leq \dfrac{1}{4n}$. (10.7.4)

Now $\operatorname*{Lt}_{n\to\infty} \dfrac{1}{4n} = 0$.

Then by (10.7.4) we get $\operatorname*{Lt}_{n\to\infty} \dfrac{\operatorname{var}(Y_1 + Y_2 + \cdots + Y_n)}{n^2} = 0$.

So the Law of Large Numbers can be applied to the sequence $\{Y_n\}$ and consequently we get

$$\frac{(Y_1 + Y_2 + \cdots + Y_n) - (p_1 + p_2 + \cdots + p_n)}{n} \xrightarrow[\text{in p}]{} 0 \text{ as } n \to \infty$$

or, $\dfrac{n\bar{X}_n - n\bar{p}}{n} \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Therefore, $\bar{X}_n - \bar{p} \xrightarrow[\text{in p}]{} 0$ as $n \to \infty$.

Ex. 12. If the sequence of random variables $X_1, X_2, \cdots, X_n$ be such that X_n is a $\chi^2(n)$ variate, then show that X_n is asymptotically normal $(n, \sqrt{2n})$.

We know that a $\chi^2(n)$ variate X_n can be expressed as

$$X_n = Y_1^2 + Y_2^2 + \cdots + Y_n^2,$$

where $Y_1, Y_2, \ldots, Y_n$ are mutually independent standard normal variates.

Again, we know that $\frac{1}{2}Y_1^2, \frac{1}{2}Y_2^2, \ldots, \frac{1}{2}Y_n^2$ are each $\gamma\left(\frac{1}{2}\right)$ variates.

Then $E\left(\frac{1}{2}Y_i^2\right) = \frac{1}{2}$ and $\operatorname{var}\left(\frac{1}{2}Y_i^2\right) = \frac{1}{2}$ for $i = 1, 2, \ldots$

So the sequence of random variables $\left\{\frac{1}{2}Y_n^2\right\}$ is such that $\frac{1}{2}Y_1^2, \frac{1}{2}Y_2^2, \ldots, \frac{1}{2}Y_n^2, \ldots$ have the same distribution with common mean $\frac{1}{2}$ and common standard deviation $\frac{1}{\sqrt{2}}$ and $\frac{1}{2}Y_1^2, \frac{1}{2}Y_2^2, \ldots, \frac{1}{2}Y_n^2$ are mutually independent for all n. Then by the Central Limit Theorem for Equal Components we get

$$\frac{\dfrac{\frac{1}{2}Y_1^2 + \frac{1}{2}Y_2^2 + \cdots + \frac{1}{2}Y_n^2}{n} - \dfrac{1}{2}}{\dfrac{\frac{1}{\sqrt{2}}}{\sqrt{n}}}$$

is asymptotically normal (0,1),

that is, $\dfrac{Y_1^2 + Y_2^2 + \cdots + Y_n^2 - n}{\sqrt{2n}}$ is asymptotically normal (0,1),

that is, $\dfrac{X_n - n}{\sqrt{2n}}$

is asymptotically normal (0,1).

So X_n is asymptotically normal $(n, \sqrt{2n})$.

Examples X

1. An urn contains 1000 white and 2000 black balls. We draw (with replacement) 300 balls. Estimate the probability that the number m of the drawn white balls satisfies the inequality $80 < m < 120$. [C. H. (Math.)'87]

[*Hint* : Let X be the random variable denoting the number m of white balls drawn in 300 drawings of balls. Let A be the event 'the ball drawn is white in a drawing'. Then $P(A) = \frac{1000}{3000} = \frac{1}{3}$ in any trial. X is then binomial $\left(300, \frac{1}{3}\right)$.

$$E(X) = 300 \times \frac{1}{3} = 100,$$

$$\sigma = \sqrt{\operatorname{var} X} = \sqrt{300 \times \frac{1}{3} \times \left(1 - \frac{1}{3}\right)} = \frac{10\sqrt{2}}{\sqrt{3}}.$$

Now $P(80 < X < 120)$

$$= P\left(-\sqrt{6} < \frac{X - 100}{\frac{10\sqrt{2}}{\sqrt{3}}} < \sqrt{6}\right)$$

$$\simeq F(\sqrt{6}) - F(-\sqrt{6}) = 2F(\sqrt{6}) - 1$$

where $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$, since by De Moivre-Laplace Limjt Thereom $\frac{X - 100}{\frac{10\sqrt{2}}{\sqrt{3}}}$ is approximately normal (0,1).

Now from the standard normal table

$$2F(\sqrt{6}) - 1 = 2(1 - 0.0071) - 1 = 0.9858.$$

Therefore, $P(80 < X < 120) \simeq 0.9858$.

Now, by Tchebycheff's inequality,

$$P(|X - 100| \geq 20) \leq \frac{200}{3 \times 400} = \frac{1}{6}.$$

Therefore, $P(80 < X < 120) = P(-20 < X - 100 < 20)$

$$= P(|X - 100| < 20) = 1 - P(|X - 100| \geq 20)$$

$$\geq 1 - \frac{1}{6} = \frac{5}{6} = 0.833$$

and this value supports the value obtained previously.]

2. If X be a random variable and $E(X^2) < \infty$, then prove that

$$P(|X| \geq a) \leq \frac{1}{a^2} E(X^2) \text{ for all } a > 0. \quad \text{[C. H. (Math.)'91]}$$

[*Hint* : Put $c = 0$, $\varepsilon = a$ in Theorem 10.2.4. The result can be obtained independently as follows :

Case I. Let X be a discrete random variable and $f_i = P(X = x_i)$, x_i being a point of the spectrum of X.

Now $|x_i| \geq a$ implies $x_i^2 \geq a^2$. So $f_i \leq \frac{x_i^2}{a^2} f_i$, since $f_i \geq 0$.

Therefore, $\sum_{|x_i| \geq a} f_i \leq \sum_{|x_i| \geq a} \frac{x_i^2}{a^2} f_i \leq \sum_i \frac{x_i^2}{a^2} f_i$, since $f_i \geq 0$ and $\frac{x_i^2}{a^2} \geq 0$.

This implies that $P(|X| \geq a) \leq \frac{1}{a^2} \sum_i x_i^2 f_i = \frac{1}{a^2} E(X^2)$ for all $a > 0$.

Case II. Let X be a continuous random variable and $f(x)$ be its probability density function.

Now $|x| \geq a$ implies $f(x) \leq \frac{x^2}{a^2} f(x)$, since $f(x) \geq 0$ for all x.

So $f(x) \leq \frac{x^2}{a^2} f(x)$ for all $x \in (-\infty, -a] \cup [a, \infty)$.

$$\text{Therefore} \int_{|x| \geq a} f(x)\, dx \leq \int_{|x| \geq a} \frac{x^2}{a^2} f(x)\, dx = \int_{-\infty}^{-a} \frac{x^2}{a^2} f(x)\, dx + \int_{a}^{\infty} \frac{x^2}{a^2} f(x)\, dx$$

$$\leq \int_{-\infty}^{-a} \frac{x^2}{a^2} f(x)\, dx + \int_{-a}^{a} \frac{x^2}{a^2} f(x)\, dx + \int_{a}^{\infty} \frac{x^2}{a^2} f(x)\, dx$$

$$\text{since} \int_{-a}^{a} \frac{x^2}{a^2} f(x)\, dx \geq 0$$

$$= \int_{-\infty}^{\infty} \frac{x^2}{a^2} f(x)\, dx.$$

Therefore $P(|X| \geq a) \leq \frac{1}{a^2} \int_{-\infty}^{\infty} x^2 f(x)\, dx = \frac{1}{a^2} E(X^2)$ for all $a > 0$.

3. Use Tchebycheff's inequality to show that for $n \geq 36$, the probability that in n throws of a fair die the number of sixes lies between $\frac{1}{6}n - \sqrt{n}$ and $\frac{1}{6}n + \sqrt{n}$ is at least $\frac{31}{36}$. [C. H. (Math.) '91]

[*Hint*: Let X be the random variable denoting the number of sixes in n throws of a fair die. Then X is a binomial $\left(n, \frac{1}{6}\right)$ variate.

$$E(X) = n \cdot \frac{1}{6}, \quad \operatorname{var} X = n \cdot \frac{1}{6}\left(1 - \frac{1}{6}\right) = \frac{5n}{36}.$$

$$\text{Now } P\left(\frac{1}{6}n - \sqrt{n} < X < \frac{1}{6}n + \sqrt{n}\right) = P\left(-\sqrt{n} < X - \frac{n}{6} < \sqrt{n}\right)$$
$$= P\left(\left|X - \frac{n}{6}\right| < \sqrt{n}\right) = 1 - P\left(\left|X - \frac{n}{6}\right| \geq \sqrt{n}\right).$$

By Tchebycheff's inequality, $P\left(\left|X - \frac{n}{6}\right| \geq \sqrt{n}\right) \leq \frac{5}{36}$.

$$\text{Therefore, } P\left(\frac{1}{6}n - \sqrt{n} < X < \frac{1}{6}n + \sqrt{n}\right) \geq 1 - \frac{5}{36} = \frac{31}{36}.$$

The number of sixes can never be negative. So $X > \frac{1}{6}n - \sqrt{n}$ will be trivially true if $\frac{n}{6} - \sqrt{n} < 0$. So we can assume that $\frac{n}{6} - \sqrt{n} \geq 0$, that is, $n \geq 36$.]

4. Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of independent random variables, where each X_i takes the values $0, \pm 2^i$. Given that

$$P(X_i = \pm 2^i) = 2^{-(2i+1)}, \qquad P(X_i = 0) = 1 - 2^{-2i}.$$

Examine whether the Law of Large Numbers holds for the sequence.

5. For the distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-m)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

determine $P\{|X - E(X)| \geq 1.5\sigma\}$. Compare the value with the limit given by Tchebycheff's inequality.

[*Hint*: $P\{|X - E(X)| \geq 1.5\sigma\}$

$$= 1 - P(m - 1.5\sigma < X < m + 1.5\sigma), \; E(X) = m$$
$$= 1 - \int_{m-1.5\sigma}^{m+1.5\sigma} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}\, dx$$
$$= 1 - \int_{-1.5}^{1.5} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}\, du, \quad \text{where } u = \frac{x-m}{\sigma}$$
$$= 1 - 2\int_{0}^{1.5} \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}\, du$$
$$= 1 - 2\left\{P(U \leq 1.5) - \frac{1}{2}\right\}$$
$$= 2\{1 - P(U \leq 1.5)\} = 2\{1 - \Phi(1.5)\}$$
$$\simeq 2(1 - 0.9332), \text{ since } \Phi(1.5) \simeq 1 - 0.0688 = 0.9332$$
$$= 0.1336.$$

Now taking $\varepsilon = 1.5\sigma > 0$, we get, by Tchebycheff's inequality, $P\{|X - m| \geq 1.5\sigma\} \leq \frac{1}{(1.5)^2} = 0.444$. We see that $0.1336 < 0.444$ which supports Tchebycheff's inequality regarding upper limit.]

6. For the random variable having geometric distribution given by

$$P(X = k) = 2^{-k}, \; k = 1, 2, 3, \ldots,$$

compute $P(|X - 2| > 2)$ and compare the result obtained by Tchebycheff's inequality.

[*Hint*: $E(X) = \sum_{k=1}^{\infty} k \cdot 2^{-k} = \frac{1}{2} + 2 \cdot \frac{1}{2^2} + 3 \cdot \frac{1}{2^3} + \cdots$

$$= \frac{1}{2}\left(1 + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2^2} + \cdots\right) = \frac{1}{2}\left(1 - \frac{1}{2}\right)^{-2} = 2.$$

$$E(X^2) = \sum_{k=1}^{\infty} k^2 \cdot 2^{-k}$$
$$= 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{2^2} + 3^2 \cdot \frac{1}{2^3} + \cdots$$

Now we have $(1 - x)^{-2} = 1 + 2x + 3x^2 + \cdots$ if $|x| < 1$

Therefore, $\frac{x}{(1-x)^2} = x + 2x^2 + 3x^3 + 4x^4 + \cdots$ if $|x| < 1$.

Here the process of term by term differentiation is valid.

Therefore, $\dfrac{d}{dx}\left\{\dfrac{x}{(1-x)^2}\right\} = 1 + 2^2 x + 3^2 x^2 + 4^2 x^3 + \cdots$ if $|x|<1$

or, $\dfrac{1}{(1-x)^2} + \dfrac{2x}{(1-x)^3} = 1 + 2^2 x + 3^2 x^2 + 4^2 x^3 + \cdots$ if $|x|<1$

or, $\dfrac{1+x}{(1-x)^3} = 1 + 2^2 x + 3^2 x^2 + 4^2 x^3 + \cdots$ if $|x|<1$.

Therefore, $\dfrac{x(1+x)}{(1-x)^3} = x + 2^2 x^2 + 3^2 x^3 + 4^2 x^4 + \cdots$ if $|x|<1$.

Now taking $x = \dfrac{1}{2}$, we get

$$\frac{1}{2} + 2^2 \cdot \frac{1}{2^2} + 3^2 \cdot \frac{1}{2^3} + \cdots = \frac{\frac{1}{2}\left(1+\frac{1}{2}\right)}{\left(1-\frac{1}{2}\right)^3} = 6.$$

Therefore, $E(X^2) = 6$ and so var $X = 6 - 4 = 2$.

Now, $P(|X-2|>2) = 1 - P(|X-2| \le 2)$

$= 1 - P(0 \le X \le 4)$

$= 1 - \{P(X=1) + P(X=2) + P(X=3) + P(X=4)\}$

$= 1 - (2^{-1} + 2^{-2} + 2^{-3} + 2^{-4}) = \frac{1}{16}$.

Again, $P(|X-2|>2) < P(|X-2| \ge 2)$,

since $P(|X-2|=2) = P(X=4) = 2^{-4} > 0$

and by Tchebycheff's inequality

$$P(|X-2| \ge 2) \le \frac{\text{var } X}{4} = \frac{1}{2}.$$

Hence, $P(|X-2|>2) < \dfrac{1}{2}$ and this is supported by $P(|X-2|>2) = \dfrac{1}{16}$ obtained before.]

7. Show that the condition in the Law of Large Numbers (W.L.L.N.) does not hold for the sequence of independent random variables $\{X_n\}_n$, where the spectrum of X_n is the set $\{-n, n, 0\}$ with $P(X_n = n) = P(X_n = -n) = \dfrac{1}{2\sqrt{n}}$, $P(X_n = 0) = 1 - \dfrac{1}{\sqrt{n}}$ for $n = 1, 2, 3, \ldots$

[*Hint* : $B_n = \text{var}(X_1 + X_2 + \cdots + X_n)$

$= 1 + 2\sqrt{2} + 3\sqrt{3} + \cdots + n\sqrt{n}$.

$$\underset{n\to\infty}{Lt} \frac{B_n}{n^2} = \underset{n\to\infty}{Lt} \frac{1 + 2\sqrt{2} + 3\sqrt{3} + \cdots + n\sqrt{n}}{n^2}$$

$$= \underset{n\to\infty}{Lt} \left\{\sqrt{n} \cdot \frac{1}{n} \sum_{r=1}^{n} \left(\frac{r}{n}\right)^{\frac{3}{2}}\right\}.$$

Now $\underset{n\to\infty}{Lt} \dfrac{1}{n} \displaystyle\sum_{r=1}^{n} \left(\frac{r}{n}\right)^{\frac{3}{2}} = \int_0^1 x^{\frac{3}{2}} dx = \dfrac{2}{5}$, which is finite and > 0.

Also $\underset{n\to\infty}{Lt} \sqrt{n} = \infty$.

So $\underset{n\to\infty}{Lt} \left\{\sqrt{n} \cdot \dfrac{1}{n} \displaystyle\sum_{r=1}^{n} \left(\frac{r}{n}\right)^{\frac{3}{2}}\right\} = \infty$,

and hence $\underset{n\to\infty}{Lt} \dfrac{B_n}{n^2} = \infty$. So the condition in the Law of Large Numbers does not hold for the given sequence of random variables.]

8. If X_n be a Poisson -n variate, then prove that X_n is asymptotically normal $(n, \sqrt{n})$.

[*Hint* : Apply Central Limit Theorem to the sequence $\{Y_n\}$ of independent random variables, where each Y_n is a Poisson -1 variate and observe that X_n can be expressed as $X_n = Y_1 + Y_2 + \cdots + Y_n$.]

9. The probability density function $f_n(x)$ of a random variable X_n is given by

$$f_n(x) = \frac{2^n \cdot x^{n-1} e^{-2x}}{\Gamma(n)} \quad \text{if } x > 0$$

$$= 0, \qquad \text{elsewhere}.$$

Show that X_n is asymptotically normal $\left(\dfrac{n}{2}, \dfrac{\sqrt{n}}{2}\right)$.

[*Hint* : Let $Y_n = 4X_n$. It can be shown that Y_n is a $\chi^2(2n)$ variate. Then by Illustrative Example 12 we find that Y_n is asymptotically normal $(2n, \sqrt{4n})$. So $\dfrac{Y_n - 2n}{\sqrt{4n}}$ is asymptotically normal $(0, 1)$.

But $$\frac{Y_n - 2n}{\sqrt{4n}} = \frac{2X_n - n}{\sqrt{n}} = \frac{X_n - \frac{n}{2}}{\frac{\sqrt{n}}{2}}.]$$

10. Let $\{X_n\}_n$ be a sequence of independent random variables, where the spectrum of X_n is the set $\{-n, n\}$ with

$$P(X_n = n) = P(X_n = -n) = \frac{1}{2}.$$

Show that the condition $\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$ in the W.L.L.N. is not satisfied for the above sequence, where $B_n = \text{var}(X_1 + X_2 + \cdots + X_n)$.

11. A symmetric die is thrown 360 times. Determine a lower bound for the probability of getting the number of sixes between 50 and 70.

Answers

4. Law of Large Numbers holds.

11. $\frac{1}{2}$.

APPENDIX

APPENDIX

SECTION A

Alternative proof of De Moivre-Laplace Limit Theorem.

X_n is binomial (n, p) variate where $p\,(0 < p < 1)$ is fixed.

Let $\phi_n(t)$ be the characteristic function of X_n.

Then $\phi_n(t) = E\left(e^{itX_n}\right) = \left(pe^{it} + q\right)^n$ for all real values of t where $q = 1 - p$ and $i = \sqrt{-1}$.

Let $X_n^* = \dfrac{X_n - np}{\sqrt{npq}}$.

We shall prove that X_n^* is asymptotically normal $(0, 1)$.

If $\phi_n^*(t)$ be the characteristic function of X_n^*, then we have

$$\phi_n^*(t) = E\left(e^{itX_n^*}\right) = E\left[e^{\frac{-itnp}{\sqrt{npq}}}\; e^{\frac{itX_n}{\sqrt{npq}}}\right]$$

$$= e^{\frac{-itnp}{\sqrt{npq}}}\; \phi_n\left(\frac{t}{\sqrt{npq}}\right)$$

$$= e^{\frac{-itnp}{\sqrt{npq}}} \left(pe^{\frac{it}{\sqrt{npq}}} + q\right)^n$$

$$= \left[e^{\frac{-it}{n}\sqrt{\frac{np}{q}}}\; pe^{\frac{it}{\sqrt{npq}}} + q \cdot e^{\frac{-it}{n}\sqrt{\frac{np}{q}}}\right]^n$$

$$= \left[pe^{\frac{it(1-p)}{\sqrt{npq}}} + q \cdot e^{\frac{-ipt}{\sqrt{npq}}}\right]^n$$

$$= \left[pe^{\frac{itq}{\sqrt{npq}}} + q \cdot e^{\frac{-ipt}{\sqrt{npq}}}\right]^n$$

Thus we get, $\phi_n^*(t) = \left[pe^{\frac{itq}{\sqrt{npq}}} + q \cdot e^{\frac{-itp}{\sqrt{npq}}}\right]^n$, ... (A)

for all real t.

Now from analysis of complex valued functions we know that for any real x,

$$e^{ix} = 1 + ix - \frac{x^2}{\lfloor 2} + \theta \frac{x^3}{\lfloor 3},$$

where θ is a complex number such that $|\theta| < 1$.

So we get

$$e^{\frac{itq}{\sqrt{npq}}} = 1 + \frac{itq}{\sqrt{npq}} - \frac{t^2q^2}{2npq} + \theta_1 \frac{t^3q^3}{6(npq)^{\frac{3}{2}}}, \quad \ldots\ldots \quad \text{(B)}$$

for some complex number θ_1 where $|\theta_1| < 1$.

$$\text{and } e^{\frac{-itp}{\sqrt{npq}}} = 1 - \frac{itp}{\sqrt{npq}} - \frac{p^2t^2}{2npq} + \theta_2 \frac{p^3t^3}{6(npq)^{\frac{3}{2}}}, \quad \ldots\ldots \quad \text{(C)}$$

for some complex number θ_2 where $|\theta_2| < 1$.

From (B) and (C) we get

$$p\,e^{\frac{itq}{\sqrt{npq}}} + q\,e^{\frac{-itp}{\sqrt{npq}}} = (p+q) - \frac{t^2}{2npq}(pq^2 + qp^2) + \frac{pqt^3(\theta_1q^2 + \theta_2p^2)}{6(npq)^{\frac{3}{2}}}$$

$$= 1 - \frac{t^2}{2n} + \frac{pq(\theta_1q^2 + \theta_2p^2)}{6} \cdot \frac{t^3}{(npq)^{\frac{3}{2}}}$$

$$= 1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}},$$

$$\text{where } \theta_3 = \frac{pq\left(\theta_1q^2 + \theta_2p^2\right)}{6}.$$

$$\text{Now } |\theta_3| = \frac{|pq|\,|\theta_1q^2 + \theta_2p^2|}{6}$$

$$\leq \frac{|pq|\left[|\theta_1|q^2 + |\theta_2|p^2\right]}{6}$$

Again $|pq| = pq < 1$ and $|\theta_1|q^2 + |\theta_2|p^2 < |\theta_1| + |\theta_2| < 1 + 1 = 2$

$(\because 0 < p < 1, 0 < q < 1)$

So, $|\theta_3| < \frac{2}{6} = \frac{1}{3} < 1.$

Thus we get

$$p\, e^{\frac{itq}{\sqrt{npq}}} + q \cdot e^{\frac{-itp}{\sqrt{npq}}} = 1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}, \qquad \text{.....} \quad \text{(D)}$$

where , $|\theta_3| < 1.$

From (A) and (D) we get

$$\phi_n^*(t) = \left[1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right]^n.$$

$$\text{So, } \log_e\left[\phi_n^*(t)\right] = n \log_e\left[1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right], \qquad \text{.....} \quad \text{(E)}$$

Again for large n we have,

$$\log_e\left[1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right]$$

$$= \left[-\frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right] - \frac{1}{2}\left[-\frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right]^2 + \cdots\cdots$$

So for large n,

$$n \log_e\left[1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right]$$

$$= -\frac{t^2}{2} + \text{ term containing } \frac{1}{\sqrt{n}} \text{ and higher powers of } \frac{1}{\sqrt{n}},$$

Here the process of taking limit term by term as $n \to \infty$ in the R.H.S. is valid.

Then, $$\underset{n\to\infty}{Lt}\; n \log_e\left[1 - \frac{t^2}{2n} + \theta_3 \frac{t^3}{(npq)^{\frac{3}{2}}}\right]$$

$$= -\frac{t^2}{2}. \qquad \text{(where } p \text{ is fixed)}$$

Hence from (E) we get

$$\underset{n\to\infty}{Lt} \log_e\left[\phi_n^*(t)\right] = -\frac{t^2}{2}$$

or,
$$\log_e\left[\underset{n\to\infty}{Lt}\, \phi_n^*(t)\right] = -\frac{t^2}{2}$$

So,
$$\underset{n\to\infty}{Lt}\, \phi_n^*(t) = e^{-\frac{t^2}{2}}$$

Now $e^{-\frac{t^2}{2}}$ is the characteristic function of a normal $(0, 1)$ variate and $\phi_n^*(t)$ is that of $X_n^* = \dfrac{X_n - np}{\sqrt{npq}}$.

Then by the limit theorem of characteristic function,

X_n^* converges in distribution to a standard normal variate as $n \to \infty$,

i.e., if $F_n^*(x)$ be the distribution function of X_n^*, then

$$\underset{n\to\infty}{Lt}\, F_n^*(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-x^2}{2}}\, dx \text{ for all } x \in R .$$

SECTION B

Additional Illustrative Examples.
(Miscellaneous)

Ex. 1. *If A, B are two events where* $P(B) = 1$ *, then prove that A, B are independent.* [C.H. (Math) 1999]

Solution : We have $AB \subseteq A$.

Then, $P(AB) \leq P(A)$ .... (1)

Again, we have

$$P(A+B) = P(A) + P(B) - P(AB)$$

or, $P(A+B) = P(A) + 1 - P(AB)$ [$\because$ here $P(B) = 1$]

So we get

$$P(AB) = 1 + P(A) - P(A+B) \quad \text{..........} \quad (2)$$

Also we have $P(A+B) \leq 1$

Then from (2) we get

$$P(AB) \geq 1 + P(A) - 1$$

$$\text{or, } P(AB) \geq P(A) \qquad \cdots\cdots\cdots \quad (3)$$

From (1) and (3) we get

$$P(AB) = P(A)$$

$$\text{or, } P(AB) = P(A) \cdot 1 = P(A) \cdot P(B)$$

Hence A, B are independent.

Ex. 2. *The probability of a person chosen at random from a population having a particular disease is* $0 \cdot 02$. *The probability that a diagnostic test gives a positive result when the disease is present is* $0 \cdot 75$ *and a negative result when the disease is not present is* $0 \cdot 97$. *Find the conditional probability of a person having the disease*

(i) when the test gives a positive result,
(ii) when the test gives a negative result.

[C. H. (Math) 1999]

Solution : Let X denote the event " the test gives a positive result." Also let A_1 denote the event "a person selected at random from the given population has the particular disease" and A_2 denote the event "the person selected at random has not the particular disease."

Here we have $A_1 + A_2 = S$ and $A_1 A_2 = 0$, where S is the certain event.

It is given that $P(A_1) = 0 \cdot 02$. Then $P(A_2) = 1 - 0 \cdot 02 = 0 \cdot 98$. Further it is given that

$$P(X|A_1) = 0 \cdot 75 \text{ and } P(\overline{X}|A_2) = 0 \cdot 97$$

Now $\qquad P(\overline{X}|A_2) = 0 \cdot 97 \Rightarrow P(X|A_2) = 1 - 0 \cdot 97 = 0 \cdot 03$

For (*i*) we are to find the value of the conditional probability $P(A_1|X)$ and for (*ii*) we are to find the value of $P(A_1|\overline{X})$.

By Bayes' theorem we get

$$P(A_1|X) = \frac{P(A_1)P(X|A_1)}{P(A_1)P(X|A_1) + P(A_2)P(X|A_2)}$$

$$= \frac{0 \cdot 02 \times 0 \cdot 75}{0 \cdot 02 \times 0 \cdot 75 + 0 \cdot 98 \times 0 \cdot 03}$$

$$= \frac{0 \cdot 015}{0 \cdot 0444} \simeq 0 \cdot 338$$

Again applying Bayes' theorem we get

$$P(A_1|\overline{X}) = \frac{P(A_1)P(\overline{X}|A_1)}{P(A_1)P(\overline{X}|A_1) + P(A_2)P(\overline{X}|A_2)}$$

$$= \frac{0 \cdot 02 \times (1 - 0 \cdot 75)}{0 \cdot 02 \times (1 - 0 \cdot 75) + 0 \cdot 98 \times 0 \cdot 97}$$

$$= \frac{0 \cdot 02 \times 0 \cdot 25}{0 \cdot 02 \times 0 \cdot 25 + 0 \cdot 98 \times 0 \cdot 97}$$

$$= \frac{0 \cdot 005}{0 \cdot 005 + 0 \cdot 9506}$$

$$= \frac{0 \cdot 005}{0 \cdot 9556} \simeq 0 \cdot 0052$$

Hence the required conditional probabilities are $0 \cdot 338$, $0 \cdot 0052$.

Ex. 3. *Consider a class $\{A, B, C\}$ of events. Suppose it is known that $\{A, B\}$ is an idependent pair and that $\{B, C\}$ is an independent pair. Does it follow that $\{A, C\}$ is an independent pair? Justify your answer.* **[C. H. (Math) 1997]**

Solution :

$\{A, B\}$ is an independent pair implies $P(AB) = P(A)\,P(B)$. $\{B, C\}$ is an independent pair implies $P(BC) = P(B)\,P(C)$.

Here we observe that

$$P(AB) = P(A)P(B) \text{ and } P(BC) = P(B)P(C)$$
$$\not\Rightarrow P(AC) = P(A)P(C)$$

We shall give one example where $P(AB) = P(A)\,P(B)$, $P(BC) = P(B)P(C)$ but $P(AC) \neq P(A)\,P(C)$.

We consider the random experiment of throwing an unbiased die.

Here the event space S is given by $S = \{1, 2, 3, 4, 5, 6\}$, where all simple events are equally likely.

Let $A = \{2, 3, 4\}$, $B = \{2, 6\}$, $C = \{2, 4, 5\}$

Here $P(A) = \frac{3}{6} = \frac{1}{2}$, $P(B) = \frac{2}{6} = \frac{1}{3}$, $P(C) = \frac{3}{6} = \frac{1}{2}$

Now $P(AB) = P(\{2\}) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(A)\, P(B)$,

$P(BC) = P(\{2\}) = \frac{1}{6} = \frac{1}{3} \cdot \frac{1}{2} = P(B)\, P(C)$.

So, $\{A, B\}$ and $\{B, C\}$ are independent pairs.

But $P(AC) = P(\{2, 4\}) = \frac{2}{6} = \frac{1}{3}$,

$P(A)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \neq \frac{1}{3}$

So here $P(AC) \neq P(A)\, P(C)$

Hence $\{A, C\}$ is not an independent pair although $\{A, B\}, \{B, C\}$ are independent pairs.

Ex. 4. *If X be a continuous random variable, then find the distribution of $F(X)$ where $F(x)$ is the distribution function of X.*

[C. H. (Math) 1997]

Solution : Let $Y = F(X)$

In real variables we have $y = F(x)$

Here $F(x)$ is monotonically non-decreasing function defined in $(-\infty, \infty)$ and X being a continuous variate $F(x)$ is continuous in $(-\infty, \infty)$.

Let y be any real number where $0 < y < 1$.

Since $F(x)$ is continuous in $(-\infty, \infty)$ and $F(-\infty) = 0$, $F(\infty) = 1$, there exists at least one real number, u, such that $F(u) = y$.

Let $S_1 = \{t : t \in R \text{ and } F(t) = y\}$

$F(u) = y \Rightarrow u \in S_1 \Rightarrow S_1 \neq \phi$

Again there exists a real number y_1 such that $y < y_1 < 1$. Due

to continuity of F, there exists a real number c such that $F(c) = y_1$. Then for any $t \in S_1$, we get $F(t) = y$ where $y < y_1 = F(c)$. So $F(t) < F(c)$.

Now F is monotonically non-decreasing in $(-\infty, \infty)$. Hence $F(t) < F(c) \Rightarrow t < c$. So it is proved that $t < c$ for all $t \in S_1$. So S_1 is a nonempty subset of R and S_1 is bounded above. Hence S_1 has a supremum, say, M belonging to R. We shall prove that $M \in S_1$. If possible let $M \notin S_1$. Then M is a limit point of S_1. Due to continuity of F at M, for every positive number ε, there exists a positive number δ such that $|F(x) - F(M)| < \varepsilon$ for all x satisfying $M - \delta < x < M + \delta$.

Now M being limit point of S_1, there exists at least one number, say, $M_1 \in S_1$ where $M_1 \neq M$ and $M - \delta < M_1 < M + \delta$. Then we get $|F(M_1) - F(M)| < \varepsilon$. But $M_1 \in S_1 \Rightarrow F(M_1) = y$. So $|y - F(M)| < \varepsilon$ where ε is arbitrary positive number.

Hence $y = F(M)$.

So $M \in S_1$ and this is contrary to the assumption, $M \notin S_1$.

Hence $M \in S_1$ and cosequently M is the greatest element of S_1. So $t > M \Rightarrow F(t) \neq y \Rightarrow F(t) > F(M)$.

Then the event $(X \le M) \Leftrightarrow [F(X) \le F(M)]$

So, $(X \le M) \Leftrightarrow (Y \le y)$, since $y = F(M)$

Hence $P(X \le M) = P(Y \le y)$

or, $F(M) = F_Y(y)$ where F_Y is the distribution function of Y.

But $F(M) = y$

So we get $F_Y(y) = y$ if $0<y<1$. Again if $y < 0$, then $(Y \le y)$ is an impossible event and so $P(Y \le y) = 0$ if $y < 0$.

Also if $y \ge 1$, then $(Y \ge y)$ is a certain event and so $P(Y \ge y) = 1$ if $y \ge 1$.

Thus we get

$$F_Y(y) = y \quad \text{if} \quad 0 < y < 1$$
$$= 0 \quad \text{if} \quad y < 0$$

$= 1 \quad \text{if} \quad y \geq 1$

Again, $F_Y(0) - F_Y(0-0) = P(Y=0)$

So, $F_Y(0) - 0 = P(Y=0) \qquad \left[\because \underset{y \to 0-}{Lt} F_Y(y) = 0\right]$

Hence $F_Y(0)) = P(Y=0) = \delta$ (say)

Let, $\delta \neq 0$

Then, $0 < \delta < 1$.

Now, $F_Y(0) \leq F_Y\left(\dfrac{\delta}{2}\right) = \dfrac{\delta}{2} < \delta = F_y(0)$

So, $F_Y(0) < F_Y(0)$ and this is absurd. Hence $F_Y(0) = 0$.

So the distribution function $F_Y(y)$ of $Y = F(X)$ is given by

$$F_Y(y) = 0 \quad \text{if} \quad y \leq 0$$
$$= y \quad \text{if} \quad 0 < y < 1$$
$$= 1 \quad \text{if} \quad y \geq 1$$

Now, if $f_Y(y)$ be the density function of Y, then $f_Y(y) = F_Y'(y)$ if $F_Y'(y)$ exists.

Then $\qquad f_Y(y) = 0 \quad \text{if} \quad y < 0$
$$= 1 \quad \text{if} \quad 0 < y < 1$$
$$= 0 \quad \text{if} \quad y > 1$$

and $f_Y(0)$, $f_Y(1)$ can be defined arbitrarily.

The form of the density function $f_Y(y)$ shows, that Y has uniform distribution in $(0, 1)$.

Ex. 5. *Suppose life (X) of a certain type of electronic tube in a given environment has a distribution function,*

$$\begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

with mean life 100 hours. Find λ. Suppose further that such a tube is observed to be operating after 80 hours. Find the distrubution function of its future life, X-80.

[C.H. (Math) 2001]

Solution:

The probability density function $f(x)$ of the random variable X is

given by $f(x) = F'(x)$ where $F'(x)$ exists.

Here $F'(x) = \lambda e^{-\lambda x}$ if $x > 0$

$= 0$ if $x < 0$

Now $f(0)$ can be defined arbitrarily. We take $f(0) = 0$

So here $f(x) = \lambda e^{-\lambda x}$ if $x > 0$

$= 0$ elsewhere.

Since $f(x) \geq 0$ for all x, we get $\lambda > 0$.

It is given that $E(X) = 100$

So,
$$\int_0^\infty x\lambda e^{-\lambda x}\,dx = 100$$

$$\lambda \underset{\beta\to\infty}{Lt} \int_0^\beta xe^{-\lambda x}dx = 100$$

or, $$\lambda \underset{\beta\to\infty}{Lt}\left[\frac{1}{\lambda^2} - \frac{\beta e^{-\lambda\beta}}{\lambda} - \frac{e^{-\lambda\beta}}{\lambda^2}\right] = 100 \qquad \ldots. \quad (A)$$

Now $\frac{1}{\lambda} \underset{\beta\to\infty}{Lt}\ e^{-\lambda\beta} = 0$ since, $\lambda > 0$

and $$\underset{\beta\to\infty}{Lt}\ \beta e^{-\lambda\beta} = \underset{\beta\to\infty}{Lt}\ \frac{\beta}{e^{\lambda\beta}}$$

$$= \underset{\beta\to\infty}{Lt}\ \frac{1}{\lambda e^{\lambda\beta}} = 0$$

So from (A) we get

$$\lambda \cdot \frac{1}{\lambda^2} = 100$$

Hence $$\lambda = \frac{1}{100}$$

Now, let $Y = X - 80$

We are to find the distribution function of Y on the hypothesis "$X > 80$", i.e., "$Y > 0$"

Let $F_Y(y/Y > 0)$ be the required conditional distribution function.

Now $F_Y(y \mid Y > 0) = P(Y \leq y \mid Y > 0)$

$$= \frac{P(Y \le y, Y > 0)}{P(Y > 0)}$$

$$= \frac{P(0 < Y \le y)}{P(Y > 0)}$$

$$= \frac{F(y) - F(0)}{1 - P(Y \le 0)}$$

$$= \frac{F(y) - F(0)}{1 - F(0)}$$

$$= \frac{1 - e^{-\lambda y} - 0}{1 - 0} \quad \text{if } y \ge 0$$

$$= 1 - e^{-\lambda y} \quad \text{if } y \ge 0$$

Also $\frac{F(y) - F(0)}{1 - F(0)} = 0$ if $y < 0$.

Hence the required conditional distribution function is given by

$$F_Y(y / Y > 0) = 1 - e^{-\lambda y} \quad \text{if } y \ge 0$$

$$= 0 \quad \text{elsewhere.}$$

Here we note that conditional distribution fuction $F_Y(y / Y > 0)$ and the given distribution function $F(x)$ are the same — this is an important property of the exponential distribution whose p.d.f. is given by $f(x) = \lambda e^{-\lambda x}$ if $x > 0$

$= 0$ elsewhere.

Ex. 6. *The random variables X and Y are both standard normal and are mutually independent. Find the expectation of* $Max\{|X|, |Y|\}$. **[C.H. (Math) 2000, 1996]**

Solution: Here the joint p.d.f $f(x,y)$ of X and Y is given by

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}}, \ -\infty < x < \infty, \ -\infty < y < \infty$$

Then $E[Max\{|X|, |Y|\}]$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Max\{|x|, |y|\} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} dx\, dy$$

Now $Max\{|x|,|y|\} = |x|$ if $|x| \geq |y|$

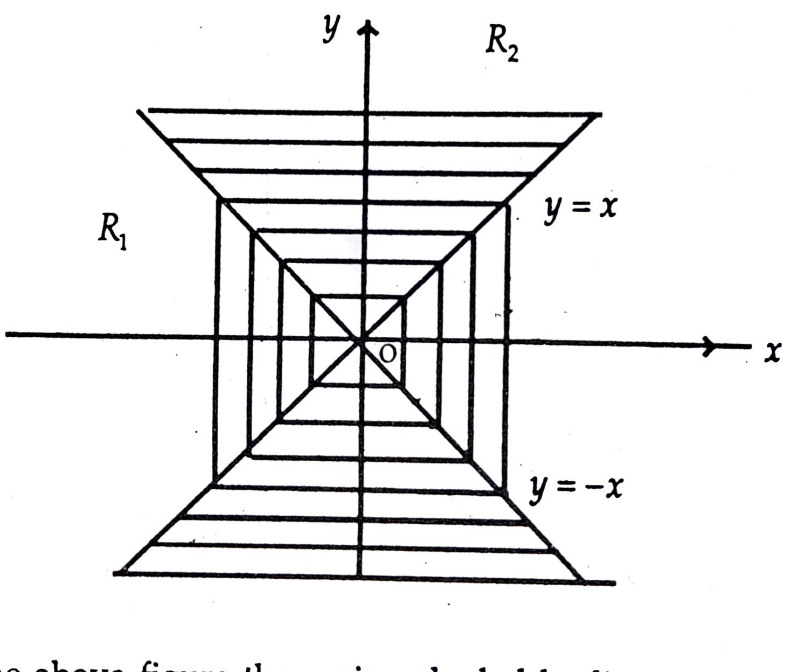
$= |y|$ if $|x| \leq |y|$

Then $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} Max\{|x|,|y|\} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} dx\, dy$

$$= \iint\limits_{R_1} Max\{|x|,|y|\} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} dx\, dy + \iint\limits_{R_2} Max\{|x|,|y|\} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} dx\, dy$$

$[R_1 = \{(x,y) : |x| \geq |y|\},\ R_2 = \{(x,y) : |x| \leq |y|\}]$

$$= \frac{1}{2\pi} \iint\limits_{R_1} |x| e^{\frac{-(x^2+y^2)}{2}} dx\, dy + \frac{1}{2\pi} \iint\limits_{R_2} |y| e^{\frac{-(x^2+y^2)}{2}} dx\, dy$$



In the above figure the region shaded by lines parallel to x-axis represents R_2 and the region shaded by lines parallel to y-axis represents R_1.

Due to symmetry we get $\iint\limits_{R_1} |x| e^{\frac{-(x^2+y^2)}{2}} dx\, dy = \iint\limits_{R_2} |y| e^{\frac{-(x^2+y^2)}{2}} dx\, dy$

Then $E[Max\{\{|X|,|Y|\}\}]$

$$= \frac{2}{2\pi} \iint_{R_2} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dx\, dy$$

$$= \frac{1}{\pi} \iint_{R_2} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dx\, dy$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{-|x|} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dy + \int_{|x|}^{\infty} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dy \right\} dx$$

Now due to symmetry we get

$$\int_{-\infty}^{-|x|} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dy = \int_{|x|}^{\infty} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dy$$

Hence we get

$$E\left[Max\{|\,X\,|, |\,Y\,|\}\right] = \frac{2}{\pi} \int_{-\infty}^{\infty} \left\{ \int_{|x|}^{\infty} |\,y\,|\, e^{\frac{-(x^2+y^2)}{2}}\, dy \right\} dx$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} \left\{ \int_{|x|}^{\infty} y\, e^{\frac{-(x^2+y^2)}{2}}\, dy \right\} dx \qquad [\because\ y \geq 0 \text{ for } y \geq |\,x\,|]$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \left\{ \int_{|x|}^{\infty} y e^{\frac{-y^2}{2}}\, dy \right\} dx$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \left\{ \int_{\frac{x^2}{2}}^{\infty} e^{-z} dz \right\} dx \qquad [\text{Putting } y^2 = 2z]$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \left[\underset{B \to \infty}{Lt} \int_{\frac{x^2}{2}}^{B} e^{-z} dz \right] dx$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \left[\underset{B \to \infty}{Lt} \left\{ -e^{-B} + e^{\frac{-x^2}{2}} \right\} \right] dx$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} e^{\frac{-x^2}{2}} \cdot e^{\frac{-x^2}{2}} \, dx = \frac{2}{\pi} \int_{-\infty}^{\infty} e^{-x^2} dx$$

Now we know that

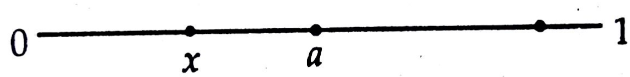
$$\int_{-\infty}^{\infty} e^{-x^2} dx = \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Hence the required expectation is given by

$$E\left[Max\{|X|, |Y|\}\right] = \frac{2}{\pi} \cdot \sqrt{\pi} = \frac{2}{\sqrt{\pi}}$$

Ex. 7. *A point a is fixed in the internal $(0, 1)$ A random variable X is uniformly distributed in that interval. Find the coefficient of correlation between the random variable X and the distance $Y (\geq 1)$ from the point a to X. Find the value of a for which the variables X and Y are uncorrelated.* **[C.H. (Math) 1996]**

Solution:



Here $Y = |X - a|$ where the p.d.f. of X is given by

$$f(x) = 1 \text{ if } 0 < x < 1$$
$$= 0 \text{ elsewhere.}$$

Now $$E(X) = \int_0^1 x \, dx = \frac{1}{2}$$

$$E(Y) = \int_0^1 |x - a| \, dx = \int_0^a |x - a| \, dx + \int_a^1 |x - a| dx$$

$$= \int_0^a (a - x) \, dx + \int_a^1 (x - a) dx$$

$$= \left[a^2 - \frac{a^2}{2} + \frac{1 - a^2}{2} - a(1 - a)\right]$$

$$= \left[\frac{1}{2} - a + a^2\right] = \frac{1 - 2a + 2a^2}{2}$$

Also $$E(XY) = \int_0^1 x|x - a| dx$$

$$=\left[\int_0^a x(a-x)\,dx+\int_a^1 x(x-a)\,dx\right]$$

$$=\left[\frac{a^3}{2}-\frac{a^3}{3}+\frac{1}{3}-\frac{a^3}{3}-\frac{a}{2}\left(1-a^2\right)\right]$$

$$=\left[\frac{a^3}{6}+\frac{1}{3}-\frac{a^3}{3}-\frac{a}{2}+\frac{a^3}{2}\right]$$

$$=\left[\frac{a^3}{3}+\frac{1}{3}-\frac{a}{2}\right]$$

$$=\left(\frac{2a^3+2-3a}{6}\right)=\frac{2a^3-3a+2}{6}$$

Then $\operatorname{cov}(X,Y)=E(XY)-E(X)\,E(Y)$

$$=\frac{2a^3-3a+2}{6}-\frac{1-2a+2a^2}{4}$$

The random variables X, Y will be uncorrelated if and only if $\operatorname{cov}(X,Y)=0$

Now $\operatorname{cov}(X,Y)=0$

$$\Rightarrow 2\left(2a^3-3a+2\right)-3\left(1-2a+2a^2\right)=0$$

$$\Rightarrow 4a^3-6a^2+1=0$$

$$\Rightarrow 4a^3-2a^2-4a^2+1=0$$

$$\Rightarrow 2a^2(2a-1)-(2a-1)(2a+1)=0$$

$$\Rightarrow (2a-1)(2a^2-2a-1)=0$$

$$\Rightarrow 2a-1=0 \text{ or } 2a^2-2a-1=0$$

Now $2a^2-2a-1=0$

$$\Rightarrow a=\frac{2\pm\sqrt{12}}{4}=\frac{1\pm\sqrt{3}}{2}$$

But $\dfrac{1+\sqrt{3}}{2}>1$ and $\dfrac{1-\sqrt{3}}{2}<0$

So, $2a^2-2a-1\neq 0$ (since $0<a<1$)

Hence $2a-1=0$

So, $a=\dfrac{1}{2}$.

Ex. 8. *How many independent Bernoulli trials are required to ensure with a probability not less then* 0·975 *the validity of the inequality* $\left|\frac{1}{n}X_n - p\right| < 0\cdot1$, *where* X_n *is the number of successes in* n *trials,* $p = \frac{1}{2}$ *is the probability of a success in one trial.* [C.H. (Math) 1997]

Solution : Here X_n has binomial (n, p) distribution.

Then $E\left(\frac{1}{n}X_n\right) = \frac{1}{n}E(X_n) = \frac{1}{n}np = p$

and $\text{var}\left(\frac{1}{n}X_n\right) = \frac{1}{n^2}\text{var}(X_n) = \frac{1}{n^2}np(1-p)$

$$= \frac{p(1-p)}{n}.$$

Now we are to find the least value of the positive integer n for which

$$P\left(\left|\frac{1}{n}X_n - p\right| < 0\cdot1\right) \not< 0\cdot975$$

or, $$P\left(\left|\frac{1}{n}X_n - p\right| < 0\cdot1\right) \geq 0\cdot975$$

or, $$P\left(\left|\frac{1}{n}X_n - p\right| \geq 0\cdot1\right) \leq 1 - 0\cdot975 = 0\cdot025 \qquad \ldots\ldots \text{(A)}$$

Now by Tchebycheff's inequality we get

$$P\left(\left|\frac{1}{n}X_n - p\right| \geq 0\cdot1\right) \leq \frac{p(1-p)}{n(0\cdot1)^2} \qquad \ldots\ldots \text{(B)}$$

Here (B) is equivalent to $P\left(\left|\frac{1}{n}X_n - p\right| \geq 0\cdot1\right) \leq \frac{1}{4n(0\cdot01)}$

or, $P\left(\left|\frac{1}{n}X_n - p\right| \geq 0\cdot1\right) \leq \frac{25}{n}$.

So, (A) will be true if

$$\frac{25}{n} \leq 0\cdot025$$

or, $\frac{1}{n} \leq \frac{1}{1000}$

or, $n \geq 1000$

Hence the required least value of n in 1000.

Ex. 9. *In a lottery with 10,000 tickets there are 100 prizes. A man buys 100 tickets. Apply Poisson approximation to binomial law to find the approximate probability of his winning at least one ticket* $\left(\frac{1}{e}=0\cdot 368\right)$ **[C.H. (Math) 1995]**

Solution : Let p be probability of getting a prize by purchasing a ticket.

Here $p = \frac{100}{10,000} = \frac{1}{100}$.

The number of tickets purchased by the man is 100.

So here $n = 100$, $p = \frac{1}{100}$. Then $np = 1$ and so if X be the random variable denoting the number of prizes won by the man, then X has approximately Poisson distribution with $\mu = 1$ [Here the exact distribution of X is binomial (n, p)]

Now the probability of winning at least one prize is

$$P(X \geq 1) = 1 - P(X = 0)$$

Here $P(X=0) \approx \frac{e^{-\mu} \cdot \mu^0}{\underline{|0}} = e^{-1}$

Hence $P(X \geq 1) \approx 1 - e^{-1} = 1 - 0\cdot 368 = 0\cdot 632$.

So the required approximate probability is 0·632.

Ex. 10. *Let* $Z_1, Z_2 \cdots Z_n \cdots$ *be a sequence of random variables. Suppose* $Z_n \to Z$ *in distribution and* $a, b \in R$. *Prove that* $a Z_n + b \to aZ + b$ *in distribution.* **[C.H. (Math) 1996 (old)]**

Solution: Let $\phi_n(t)$ be the characteristic function of Z_n $(n = 1, 2, \cdots\cdots)$ and $\chi(t)$ be that of Z.

Then by the limit theorem of characteristic function we find that $Z_n \to Z$ in distribution $\Rightarrow \underset{n\to\infty}{Lt}\ \phi_n(t) = \chi(t)$.

If $\psi_n(t)$ be the characteristic function of $aZ_n + b$, we see that

$\psi_n(t) = E\left[e^{it(aZ_n+b)}\right] = e^{itb}\, E\left[e^{itaZ_n}\right]$

Now $\phi_n(t) = E\left[e^{itZ_n}\right]$ for all real t.

So, $\phi_n(at) = E\left[e^{iatZ_n}\right]$

Hence, $\psi_n(t) = e^{itb}\phi_n(at)$

Here, $\underset{n\to\infty}{Lt}\ \phi_n(t) = \chi(t)$ for all real t.

So, $\underset{n\to\infty}{Lt}\ \psi_n(t) = e^{itb}\ \underset{n\to\infty}{Lt}\ \phi_n(at) = e^{itb}\ \chi(at)$ ..... (A)

Now let $\psi(t)$ be the characteristic function of $aZ + b$.

Then $\psi(t) = E\left[e^{it(aZ+b)}\right]$

or, $\psi(t) = e^{itb} E\left[e^{iatZ}\right]$ ..... (B)

Again, $\chi(t) = E\left[e^{itZ}\right]$ for all real t.

So, $\chi(at) = E\left[e^{iatZ}\right]$ ..... (C)

Hence, using (A) and (C) we get

$$\underset{n\to\infty}{Lt}\ \psi_n(t) = e^{itb} E\left[e^{iatZ}\right]$$

Then using (B) we get $\underset{n\to\infty}{Lt}\ \psi_n(t) = \psi(t)$, where we remember that $\psi_n(t)$ is the characteristic function of $aZ_n + b$ and $\psi(t)$ is that of $aZ + b$.

So by the limit theorem of characteristic function we get,

$$aZ_n + b \to aZ + b \text{ in distribution.}$$

Ex. 11. *If the random variable X is normal $(\lambda, 1)$, show that $E(Y) = \frac{1}{\lambda}$ $(\lambda > 0)$ where $Y = \frac{[1-\Phi(X)]}{\phi(X)}$; Φ and ϕ are respectively the distribution function and density function of a standard normal variate.*

[C.H. (Math) 1999]

Solution : The probability density function $f(x)$ of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\lambda)^2}{2}}, \ -\infty < x < \infty.$$

Also we know that $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ and $\Phi'(x) = \phi(x)$ for all real values of x.

$$\text{So, } E(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{[1-\Phi(x)]}{\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}} e^{\frac{-(x-\lambda)^2}{2}} dx$$

$$= \int_{-\infty}^{\infty} [1-\Phi(x)] e^{\frac{-(\lambda^2-2\lambda x)}{2}} dx$$

$$= e^{\frac{-\lambda^2}{2}} \int_{-\infty}^{\infty} [1-\Phi(x)] e^{\lambda x} dx$$

$$= e^{\frac{-\lambda^2}{2}} \underset{\substack{B_1\to-\infty\\ B_2\to\infty}}{Lt} \int_{B_1}^{B_2} [1-\Phi(x)] e^{\lambda x} dx$$

$$= e^{\frac{-\lambda^2}{2}} \underset{\substack{B_1\to-\infty\\ B_2\to\infty}}{Lt} \left[\frac{e^{\lambda x}\{1-\Phi(x)\}}{\lambda}\right]_{B_1}^{B_2} - e^{\frac{-\lambda^2}{2}} \underset{\substack{B_1\to-\infty\\ B_2\to\infty}}{Lt} \int_{B_1}^{B_2} -\frac{\Phi'(x)e^{\lambda x}}{\lambda} dx$$

$$= e^{\frac{-\lambda^2}{2}} \underset{\substack{B_1\to-\infty\\ B_2\to\infty}}{Lt} \left[\frac{e^{\lambda B_2}\{1-\Phi(B_2)\}}{\lambda} - \frac{e^{\lambda B_1}\{1-\Phi(B_1)\}}{\lambda}\right]$$

$$+ e^{\frac{-\lambda^2}{2}} \underset{\substack{B_1\to-\infty\\ B_2\to\infty}}{Lt} \int_{B_1}^{B_2} \frac{1}{\lambda} e^{\lambda x} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

Now $\quad \underset{B_1\to-\infty}{Lt} \left[e^{\lambda B_1}\{1-\Phi(B_1)\}\right] = 0$, Since $\underset{B_1\to-\infty}{Lt} \Phi(B_1) = 0$

and $\underset{B_1\to-\infty}{Lt} e^{\lambda B_1} = 0$ for $\lambda > 0$.

Also, $\quad \underset{B_2\to\infty}{Lt} \left[e^{\lambda B_2}\{1-\Phi(B_2)\}\right]$

$$= \underset{B_2\to\infty}{Lt} \frac{1-\Phi(B_2)}{e^{-\lambda B_2}}$$ [Form $\frac{0}{0}$, since $\underset{B_2\to\infty}{Lt} \Phi(B_2) = 1$

and $\underset{B_2\to\infty}{Lt} e^{-\lambda B_2} = 0$ for $\lambda > 0$]

$$= \underset{B_2\to\infty}{Lt} \frac{-\phi'(B_2)}{-\lambda\, e^{-\lambda B_2}}$$

$$= \underset{B_2\to\infty}{Lt} \frac{1}{\lambda} e^{\lambda B_2} \frac{1}{\sqrt{2\pi}} e^{\frac{-B_2^2}{2}}$$

$$= \underset{B_2\to\infty}{Lt} \frac{1}{\sqrt{2\pi}\,\lambda} e^{\frac{-(B_2-\lambda)^2}{2}} \cdot e^{\frac{\lambda^2}{2}}$$

$$= 0 .$$

So, we get

$$E(Y) = e^{\frac{-\lambda^2}{2}} \cdot 0 + \frac{1}{\lambda} e^{\frac{-\lambda^2}{2}} \cdot \underset{\substack{B_1\to-\infty \\ B_2\to\infty}}{Lt} \int_{B_1}^{B_2} \frac{e^{\left(\lambda x - \frac{x^2}{2}\right)}}{\sqrt{2\pi}}\, dx$$

$$= \frac{1}{\lambda} \underset{\substack{B_1\to-\infty \\ B_2\to\infty}}{Lt} \frac{1}{\sqrt{2\pi}} \int_{B_1}^{B_2} e^{-\frac{(x-\lambda)^2}{2}}\, dx$$

$$= \frac{1}{\lambda} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\lambda)^2}{2}}\, dx = \frac{1}{\lambda} \cdot 1 = \frac{1}{\lambda}$$

$\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\lambda)^2}{2}} \right.$ being the p.d.f. of a normal $(\lambda, 1)$ variate,

$\left. \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x-\lambda)}{2}}\, dx = 1 \right]$

Hence it is proved that

$$E(Y) = \frac{1}{\lambda} .$$

Ex. 12. *In a Bernoulli sequence of n trials it is known that there are exactly r successes. Find the conditional probability of a success on the i-th trial.* **[C. H. (Math) 1996]**

Solution : Let p be the probability of success in each trial.
Let A_r denote the event " r successes in a Bernoulli sequence of n trials " and X denote the event " a success on the i-th trial."
We are to find the value of the conditional probability $P(X|A_r)$.
From the definition of the conditional probability we find that

$$P(X|A_r) = \frac{P(XA_r)}{P(A_r)}$$

We see that the event XA_r happens if and only if X and A_r simultaneously happen, *i.e.*, if and only if the event "a success in the i-th trial and exactly $r-1$ successes in the remaining $n-1$ trials" happens.

Now by binomial law we find that $P(A_r) = {}^{n}c_r p^r(1-p)^{n-r}$ and the probability of getting exactly $r-1$ successes in a Bernoulli sequence of $n-1$ trials is ${}^{n-1}c_{r-1} p^{r-1}(1-p)^{(n-1)-(r-1)}$

$${}^{n-1}c_{r-1} p^{r-1}(1-p)^{n-r}$$

Then, since the trials are independent, we get

$$P(XA_r) = p \cdot {}^{n-1}c_{r-1} \cdot p^{r-1}(1-p)^{n-r}$$

Hence,
$$P(X|A_r) = \frac{{}^{n-1}c_{r-1} p^r (1-p)^{n-r}}{{}^{n}c_r p^r(1-p)^{n-r}}$$

$$= \frac{{}^{n-1}c_{r-1}}{{}^{n}c_r}$$

$$= \frac{\underline{|n-1}}{\underline{|r-1}\ \underline{|n-r}} \cdot \frac{\underline{|r}\ \underline{|n-r}}{\underline{|n}}$$

$$= \frac{r}{n}.$$

So, the required conditional probability is $\frac{r}{n}$.

Ex. 13. *Three marksmen can hit the target with probabilities $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}$ respectively. They shoot simultaneously and there are two hits. Find the probability of missing the target by each of the three marksmen.* **[C.H. (Math) 1997]**

Solution : We denote the events of hitting the target by the three marksmen respectively by A, B, C. It is given that

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{2}{3}, \quad P(C) = \frac{3}{4}.$$

Let X denote the event "there are two hits when the three marksmen shoot simultaneously".

Here we are to find the values of the conditional probabilities

$$P(\overline{A}|X),\ P(\overline{B}|X),\ P(\overline{C}|X).$$

We find that the event X can be expressed as

$$X = AB\overline{C} + BC\overline{A} + CA\overline{B}$$

where $AB\overline{C}$, $BC\overline{A}$, $CA\overline{B}$ are pairwise mutually exclusive events.

Then $P(X) = P(AB\overline{C}) + P(BC\overline{A}) + P(CA\overline{B})$.

$$= P(A)\,P(B)\,P(\overline{C}) + P(B)\,P(C)\,P(\overline{A}) + P(C)\,P(A)\,P(\overline{B})$$

[$\because$ A, B, C are independent]

$$= \frac{1}{2}\cdot\frac{2}{3}\left(1-\frac{3}{4}\right) + \frac{2}{3}\cdot\frac{3}{4}\left(1-\frac{1}{2}\right) + \frac{3}{4}\cdot\frac{1}{2}\left(1-\frac{2}{3}\right)$$

$$= \frac{1}{12} + \frac{1}{4} + \frac{1}{8} = \frac{11}{24}$$

Now $$P(\overline{A}|X) = \frac{P(\overline{A}X)}{P(X)}$$

$$= \frac{P\left[\overline{A}\left(AB\overline{C} + BC\overline{A} + CA\overline{B}\right)\right]}{P(X)}$$

$$= \frac{P\left(BC\overline{A}\right)}{P(X)} = \frac{P(B)\,P(C)\,P(\overline{A})}{P(X)}$$

$$= \frac{\frac{2}{3}\cdot\frac{3}{4}\cdot\frac{1}{2}}{\frac{11}{24}} = \frac{1}{4}\times\frac{24}{11} = \frac{6}{11}$$

$$P(\overline{B}|X) = \frac{P(\overline{B}X)}{P(X)} = \frac{P\left(AC\overline{B}\right)}{P(X)}$$

$$= \frac{\frac{1}{2}\cdot\frac{3}{4}\cdot\left(1-\frac{2}{3}\right)}{\frac{11}{24}}$$

$$= \frac{1}{8}\times\frac{24}{11} = \frac{3}{11}$$

$$P(\overline{C}|X)=\frac{P(\overline{C}X)}{P(X)}=\frac{P(AB\overline{C})}{P(X)}$$

$$=\frac{\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{4}}{\frac{11}{24}}=\frac{2}{11}.$$

So the required conditional probabilities are respectively $\frac{6}{11},\frac{3}{11},\frac{2}{11}$.

Ex. 14. *The numbers X, Y are chosen at random from a set of natural numbers $\{1,2,\cdots N\}$, $N\geq 3$ with replacement. Find the probability that $|X^2-Y^2|$ is divisible by 3.*

Solution : Here any value from the set $\{1,2,\cdots N\}$ can be taken by each of X and Y independently. So the total number of different ordered pairs (X,Y) which can be chosen from the set $\{1,2,\cdots N\}$ with replacement is equal to ${}^NC_1\times{}^NC_1=N^2$.

Let A denote the event " $|X^2-Y^2|$ is divisible by 3".

We know that any natural number is of the form $3k$ or $3k+1$ or $3k+2$ where k is a non-negative integer.

Now " $|X^2-Y^2|$ is divisible by 3" happens if and only if one of the following pairwise mutually exclusive events happens :

A_1 "X and Y are both of the form $3k$"

A_2 "X and Y are both of the form $3k+1$"

A_3 "X and Y are both of the form $3k+2$"

A_4 "X is of the form $3k+1$ and Y is of the form $3k+2$"

A_5 "X is of the form $3k+2$ *and* Y is of the form $3k+1$"

Now the number of numbers of the form $3k$ *i.e.* the number of numbers which are divisible by 3 in the set $\{1,2,\cdots N\}$ is equal to $\left[\frac{N}{3}\right]$ which denotes the greatest integer not greater than $\frac{N}{3}$.

So the number of outcomes favourable to the event A_1 is $\left[\frac{N}{3}\right]^2$

Now let the number of numbers of the form $3k+1$ be x and the number of numbers of the form $3k+2$ be y in the set $\{1, 2, \cdots N\}$.

Then $x+y+\left[\frac{N}{3}\right]=N$ and the number of outcomes favourable to A_2, A_3, A_4, A_5 are respectively $x^2,\ y^2,\ xy,\ yx$.

Thus we find that the number of outcomes favourable to the required event A is equal to $\left[\frac{N}{3}\right]^2+x^2+y^2+2xy$

$$=\left[\frac{N}{3}\right]^2+(x+y)^2$$

$$=\left[\frac{N}{3}\right]^2+\left(N-\left[\frac{N}{3}\right]\right)^2$$

Then assuming that all simple events are equally likely, by the classical definition we get,

$$P(A)=\frac{\left[\frac{N}{3}\right]^2+\left(N-\left[\frac{N}{3}\right]\right)^2}{N^2}$$

which is the required probability.

Ex. 15. *There are three persons aged 50 yrs, 60 yrs and 70 yrs old respectively. The probability to live 10 years more is $\frac{4}{5}$ for a 50 yrs old, $\frac{1}{2}$ for a 60 yrs old and $\frac{1}{5}$ for a 70 yrs old person. Find the probability that at least two of them will survive 10 yrs more.* **[C.H. (Math) -1998]**

Solution : Let A, B, C denote respectively the events " 50 yrs old person will survive 10 years more, " " 60 yrs old person will survive 10

yrs more" "70 yrs old person will survive 10 yrs more".

Then the event "at least two of them will survive 10 yrs more" is expressed by $AB\overline{C} + BC\overline{A} + CA\overline{B} + ABC$.

Here we can assume that the events A, B, C are mutually independent. Then the required probability is equal to

$$P(AB\overline{C} + BC\overline{A} + CA\overline{B} + ABC)$$

$$= P(AB\overline{C}) + P(BC\overline{A}) + P(CA\overline{B}) + P(ABC)$$

[$\because$ $AB\overline{C}, BC\overline{A}, CA\overline{B}, ABC$ are pairwise mutually execlusive events.]

$$= P(A)P(B)P(\overline{C}) + P(B)P(C)P(\overline{A}) + P(C)P(A)P(\overline{B}) + P(A)P(B)P(C)$$

since A, B, C are mutually independent.

It is given that $P(A) = \frac{4}{5}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{5}$.

Then the required probability is equal to

$$\frac{4}{5}\cdot\frac{1}{2}\cdot\frac{4}{5} + \frac{1}{2}\cdot\frac{1}{5}\cdot\frac{1}{5} + \frac{1}{5}\cdot\frac{4}{5}\cdot\frac{1}{2} + \frac{4}{5}\cdot\frac{1}{2}\cdot\frac{1}{5}$$

$$= \frac{8}{25} + \frac{1}{50} + \frac{2}{25} + \frac{2}{25}$$

$$= \frac{25}{50} = \frac{1}{2}.$$

Hence the probability that at least two of the given persons will survive 10 years more is equal to $\frac{1}{2}$.

Ex. 16. *In a repeated throw of two dice what is the probability that 6 appears first time at atleast one of the dice in k-th throw?*

[C.H. (Math)(Old) -1996]

Solution: Let A denote the event "6 appears at atleast one of the two dice " in any throw of two dice.

Then we have a sequence of Bernoulli trials where the event A is taken as success.

Now the outcomes favourable to the event A are (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), (1,6), (2,6), (3,6), (4,6), (5,6)

We see that the number of outcomes favourable to A is 11.

Also the total number of possible outcomes in any throw of two dice is ${}^6C_1 \times {}^6C_1 = 36$. Then, assuming that all the outcomes of a throw are equally likely, the probability of success in any throw is given by $P(A) = \frac{11}{36}$.

Let X denote the event "6 appears first time at atleast one of the dice in k-th throw."

Then X happens if and only if we get $(k-1)$ failures in the first $k-1$ trials and a success in the k-th trial of above mentioned sequence of Bernoulli trials, where the probability of failure in any trial is equal to $P(\overline{A}) = 1 - \frac{11}{36} = \frac{25}{36}$.

Since the trials are independent, the required probability $P(X)$ is given by

$$P(X) = \left(\frac{25}{36}\right)^{k-1} \cdot \frac{11}{36}$$

$$= \frac{(25)^{k-1} \cdot 11}{(36)^k}.$$

SECTION C

TABLES

Table I. Standard normal distribution.

Here values of $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$ are given for different values of x.

Here, we note that,

$$\Phi(-x) = 1 - \Phi(x).$$

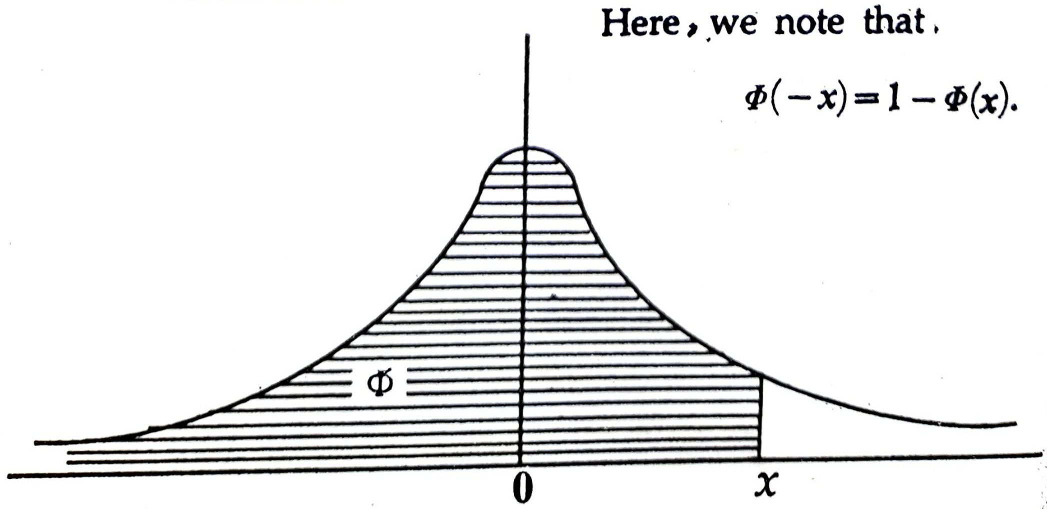


Table II. χ^2 distribution.

The table gives the values of $\chi^2_{\varepsilon, n}$ for different values of ε $(0 < \varepsilon < 1)$ and the number of degrees of freedom n where $P(\chi^2 > \chi^2_{\varepsilon, n}) = \varepsilon$.

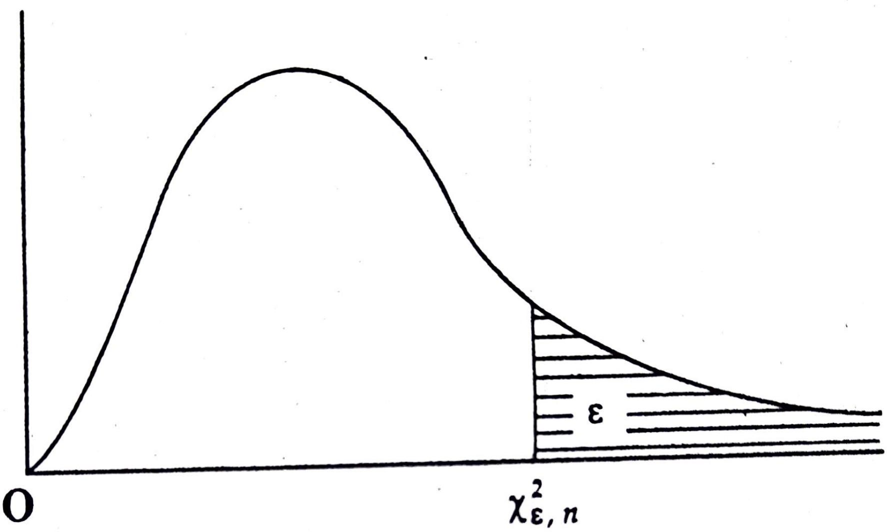


Table III. t - distribution.

The table gives the values of $t_{\varepsilon, n}$ for different values of ε and the number of degrees of freedom n where $P(t > t_{\varepsilon, n}) = \varepsilon$ $(0 < \varepsilon < 1)$.

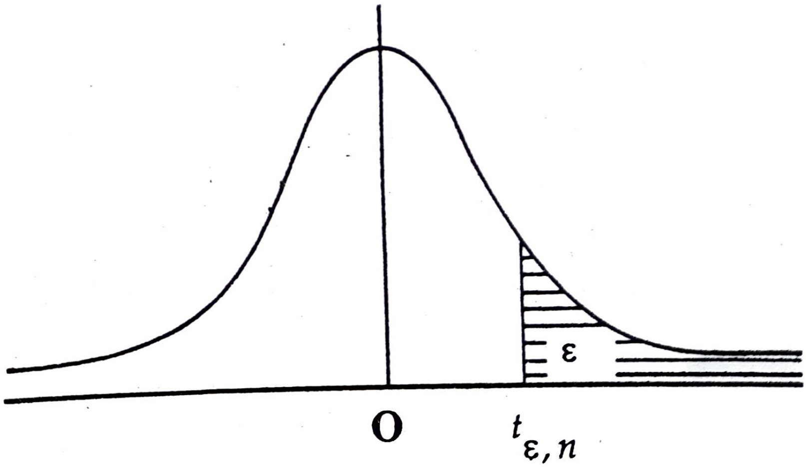
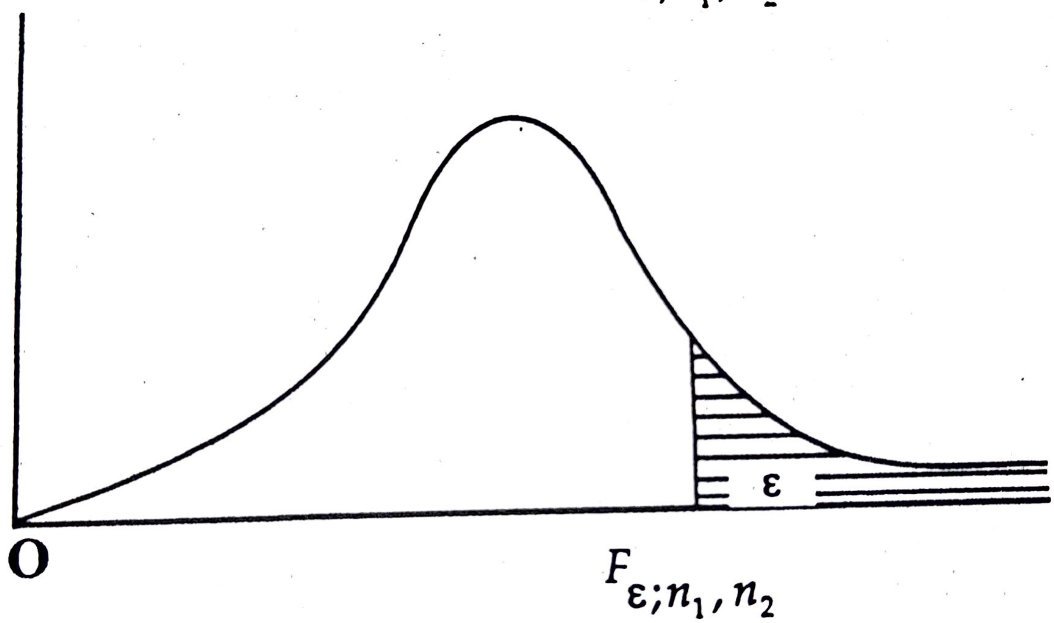


Table IV. F-distribution.

The table IV(A) gives the values of $F_{\varepsilon;n_1,n_2}$ for $\varepsilon = 0{\cdot}05$ and the table IV(B) gives the values of $F_{\varepsilon;n_1,n_2}$ for $\varepsilon = 0{\cdot}01$ (for different values of n_1, n_2) where $P(F > F_{\varepsilon;n_1,n_2}) = \varepsilon$.



TABLES

Table I
Standard Normal Distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000000	0.30	0.6179114	0.60	0.7257469
0.01	0.5039894	0.31	0.6217195	0.61	0.7290691
0.02	0.5079783	0.32	0.6255158	0.62	0.7323711
0.03	0.5119665	0.33	0.6293000	0.63	0.7356527
0.04	0.5159534	0.34	0.6330717	0.64	0.7389137
0.05	0.5199388	0.35	0.6368307	0.65	0.7421539
0.06	0.5239222	0.36	0.6405764	0.66	0.7453731
0.07	0.5279032	0.37	0.6443088	0.67	0.7485711
0.08	0.5318814	0.38	0.6480273	0.68	0.7517478
0.09	0.5358564	0.39	0.6517317	0.69	0.7549029
0.10	0.5398278	0.40	0.6554217	0.70	0.7580363
0.11	0.5437953	0.41	0.6590970	0.71	0.7611479
0.12	0.5477584	0.42	0.6627573	0.72	0.7642375
0.13	0.5517168	0.43	0.6664022	0.73	0.7673049
0.14	0.5556700	0.44	0.6700314	0.74	0.7703500
0.15	0.5596177	0.45	0.6736448	0.75	0.7733726
0.16	0.5635595	0.46	0.6772419	0.76	0.7763727
0.17	0.5674949	0.47	0.6808225	0.77	0.7793501
0.18	0.5714237	0.48	0.6843863	0.78	0.7823046
0.19	0.5753454	0.49	0.6879331	0.79	0.7852361
0.20	0.5792597	0.50	0.6914625	0.80	0.7881446
0.21	0.5831662	0.51	0.6949743	0.81	0.7910299
0.22	0.5870644	0.52	0.6984682	0.82	0.7938919
0.23	0.5909541	0.53	0.7019440	0.83	0.7967306
0.24	0.5948349	0.54	0.7054015	0.84	0.7995458
0.25	0.5987063	0.55	0.7088403	0.85	0.8023375
0.26	0.6025681	0.56	0.7122603	0.86	0.8051055
0.27	0.6064199	0.57	0.7156612	0.87	0.8078498
0.28	0.6102612	0.58	0.7190427	0.88	0.8105703
0.29	0.6140919	0.59	0.7224047	0.89	0.8132671
0.30	0.6179114	0.60	0.7257469	0.90	0.8159399

Table I (*continued*)
Standard Normal Distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.90	0.8159399	1.20	0.8849303	1.50	0.9331928
0.91	0.8185887	1.21	0.8868606	1.51	0.9344783
0.92	0.8212136	1.22	0.8887676	1.52	0.9357445
0.93	0.8238145	1.23	0.8906514	1.53	0.9369916
0.94	0.8263912	1.24	0.8925123	1.54	0.9382198
0.95	0.8289439	1.25	0.8943502	1.55	0.9394292
0.96	0.8314724	1.26	0.8961653	1.56	0.9406201
0.97	0.8339768	1.27	0.8979577	1.57	0.9417924
0.98	0.8364569	1.28	0.8997274	1.58	0.9429466
0.99	0.8389129	1.29	0.9014747	1.59	0.9440826
1.00	0.8413447	1.30	0.9031995	1.60	0.9452007
1.01	0.8437524	1.31	0.9049021	1.61	0.9463011
1.02	0.8461358	1.32	0.9065825	1.62	0.9473839
1.03	0.8484950	1.33	0.9082409	1.63	0.9484493
1.04	0.8508300	1.34	0.9298773	1.64	0.9494974
1.05	0.8531409	1.35	0.9114920	1.65	0.9505285
1.06	0.8554277	1.36	0.9130850	1.66	0.9515428
1.07	0.8576903	1.37	0.9146565	1.67	0.9525403
1.08	0.8599289	1.38	0.9162067	1.68	0.9535213
1.09	0.8621834	1.39	0.9177356	1.69	0.9544860
1.10	0.8643339	1.40	0.9192433	1.70	0.9554345
1.11	0.8665005	1.41	0.9207302	1.71	0.9563671
1.12	0.8686431	1.42	0.9221962	1.72	0.9572838
1.13	0.8707619	1.43	0.9236415	1.73	0.9581849
1.14	0.8728568	1.44	0.9250663	1.74	0.9590705
1.15	0.8749281	1.45	0.9264707	1.75	0.9599408
1.16	0.8769756	4.46	0.9278550	1.76	0.9607961
1.17	0.8789995	1.47	0.9292191	1.77	0.9616364
1.18	0.8809999	1.48	0.9305634	1.78	0.9624620
1.19	0.8829768	1.49	0.9318879	1.79	0.9632730
1.20	0.8849303	1.50	0.9331928	1.80	0.9640697

Table I (*continued*)
Standard Normal Distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1.80	0.9640697	2.10	0.9821356	2.40	0.9918025
1.81	0.9648521	2.11	0.9825708	2.41	0.9920237
1.82	0.9656205	2.12	0.9829970	2.42	0.9922397
1.83	0.9663750	2.13	0.9834142	2.43	0.9924506
1.84	0.9671159	2.14	0.9838226	2.44	0.9926564
1.85	0.9678432	2.15	0.9842224	2.45	0.9928572
1.86	0.9685572	2.16	0.9846137	2.46	0.9930531
1.87	0.9692581	2.17	0.9849966	2.47	0.9932443
1.88	0.9699460	2.18	0.9853713	2.48	0.9934309
1.89	0.9706210	2.19	0.9857379	2.49	0.9936128
1.90	0.9712834	2.20	0.9860966	2.50	0.9937903
1.91	0.9719334	2.21	0.9864474	2.51	0.9939634
1.92	0.9725711	2.22	0.9867906	2.52	0.9941323
1.93	0.9731966	2.23	0.9871263	2.53	0.9942969
1.94	0.9738102	2.24	0.9874545	2.54	0.9944574
1.95	0.9744119	2.25	0.9877755	2.55	0.9946139
1.96	0.9750021	2.26	0.9880894	2.56	0.9947664
1.97	0.9755808	2.27	0.9883962	2.57	0.9949159
1.98	0.9761482	2.28	0.9886962	2.58	0.9950600
1.99	0.9767045	2.29	0.9889893	2.59	0.9952012
2.00	0.9772499	2.30	0.9892759	2.60	0.9953388
2.01	0.9777844	2.31	0.9895559	2.61	0.9954729
2.02	0.9783083	2.32	0.9898296	2.62	0.9956035
2.03	0.9788217	2.33	0.9900969	2.63	0.9957308
2.04	0.9793248	2.34	0.9903581	2.64	0.9958547
2.05	0.9798178	2.35	0.9906133	2.65	0.9959754
2.06	0.9803007	2.36	0.9908625	2.66	0.9960930
2.07	0.9807738	2.37	0.9911060	2.67	0.9962074
2.08	0.9812372	2.38	0.9913437	2.68	0.9963189
2.09	0.9816911	2.39	0.9915758	2.69	0.9964274
2.10	0.9821356	2.40	0.9918025	2.70	0.9965330

Table I (*continued*)
Standard Normal Distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
2.70	0.9965330	3.00	0.9986501	3.30	0.9995166
2.71	0.9966358	3.01	0.9986938	3.31	0.9995335
2.72	0.9967359	3.02	0.9987361	3.32	0.9995499
2.73	0.9968333	3.03	0.9987772	3.33	0.9995658
2.74	0.9969280	3.04	0.9988171	3.34	0.9995811
2.75	0.9970202	3.05	0.9988558	3.35	0.9995959
2.76	0.9971099	3.06	0.9988933	3.36	0.9996103
2.77	0.9971972	3.07	0.9989297	3.37	0.9996242
2.78	0.9972821	3.08	0.9989650	3.38	0.9996376
2.79	0.9973646	3.09	0.9989992	3.39	0.9996505
2.80	0.9974449	3.10	0.9990324	3.40	0.9996631
2.81	0.9975229	3.11	0.9990646	3.41	0.9996752
2.82	0.9975988	3.12	0.9990957	3.42	0.9996869
2.83	0.9976726	3.13	0.9991260	3.43	0.9996982
2.84	0.9977443	3.14	0.9991553	3.44	0.9997091
2.85	0.9978140	3.15	0.9991836	3.45	0.9997197
2.86	0.9978818	3.16	0.9992112	3.46	0.9997299
2.87	0.9979476	3.17	0.9992378	3.47	0.9997398
2.88	0.9980116	3.18	0.9992636	3.48	0.9997493
2.89	0.9980738	3.19	0.9992886	3.49	0.9997585
2.90	0.9981342	3.20	0.9993129	3.50	0.9997674
2.91	0.9981929	3.21	0.9993363	3.51	0.9997759
2.92	0.9982498	3.22	0.9993590	3.52	0.9997842
2.93	0.9983052	3.23	0.9993810	3.53	0.9997922
2.94	0.9983589	3.24	0.9994024	3.54	0.9997999
2.95	0.9984111	3.25	0.9994230	3.55	0.9998074
2.96	0.9984618	3.26	0.9994429	3.56	0.9998146
2.97	0.9985110	3.27	0.9994623	3.57	0.9998215
2.98	0.9985110	3.28	0.9994810	3.58	0.9998282
2.99	0.9986051	3.29	0.9994991	3.59	0.9998347
3.00	0.9986501	3.30	0.9995166	3.60	0.9998409

Table I (*continued*)
Standard Normal Distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
3.60	0.9998409	3.75	0.9999116	3.90	0.9999519
3.61	0.9998469	3.76	0.9999150	3.91	0.9999539
3.62	0.9998527	3.77	0.9999184	3.92	0.9999557
3.63	0.9998583	3.78	0.9999216	3.93	0.9999575
3.64	0.9998637	3.79	0.9999247	3.94	0.9999593
3.65	0.9998689	3.80	0.9999277	3.95	0.9999609
3.66	0.9998739	3.81	0.9999305	3.96	0.9999625
3.67	0.9998787	3.82	0.9999333	3.97	0.9999641
3.68	0.9998834	3.83	0.9999359	3.98	0.9999655
3.69	0.9998879	3.84	0.9999385	3.99	0.9999670
3.70	0.9998922	3.85	0.9999409	4.00	0.9999683
3.71	0.9998964	3.86	0.9999433		
3.72	0.9999004	3.87	0.9999456		
3.73	0.9999043	3.88	0.9999478		
3.74	0.9999080	3.89	0.9999499		
3.75	0.9999116	3.90	0.9999519		

Table II

χ^2 – Distribution

Values of $\chi^2_{\varepsilon, n}$

n/ε	0.995	0.950	0.050	0.025	0.010	0.005	0.001
1	392704×10^{-10}	393214×10^{-8}	3.84146	5.02389	6.63490	787944	10.828
2	0.0100251	0.102587	5.99147	7.37776	9.21034	10.5966	13.816
3	0.0717212	0.351846	7.81473	9.34840	11.3449	12.8381	16.266
4	0.206990	0.710721	9.48733	11.1433	13.2767	14.8602	18.467
5	0.411740	1.145476	11.0705	12.8325	15.0863	16.7496	20.515
6	0.675727	1.63539	12.5916	14.4494	16.8119	18.5476	22.458
7	0.989265	2.16735	14.0671	16.0128	18.4753	20.2777	24.322
8	1.344419	2.73264	15.5073	17.5346	20.0902	21.9550	26.125
9	1.734926	3.32511	16.9190	19.0228	21.6660	23.5893	27.877
10	2.15585	3.94030	18.3070	20.4831	23.2093	25.1882	29.588
11	2.60321	4.57481	19.6751	21.9200	24.7250	26.7569	31.264
12	3.07382	5.22603	21.0261	23.3367	26.2170	28.2995	32.909
13	3.56503	5.89186	22.3621	24.7356	27.6883	29.8194	34.528
14	4.07468	6.57063	23.6848	26.1190	29.1413	31.3193	36.123
15	4.60094	7.26094	24.9958	27.4884	30.5779	32.8013	37.697
16	5.14224	7.96164	26.2962	28.8454	31.9999	34.2672	39.252
17	5.69724	8.67176	27.5871	30.1910	33.4087	35.7185	40.790
18	6.26481	9.39046	28.8693	31.5264	34.8053	37.1564	42.312
19	6.84398	10.1170	30.1435	32.8523	36.1908	38.5822	43.820
20	7.43386	10.8508	31.4104	34.1696	37.5662	39.9968	45.315
21	8.03366	11.5913	32.6705	35.4789	38.9321	41.4010	46.797
22	8.64272	12.3380	33.9244	36.7807	40.2894	42.7956	48.268
23	9.26042	13.0905	35.1725	38.0757	41.6384	44.1813	49.728
24	9.88623	13.8484	36.4151	39.3641	42.9798	45.5585	51.179
25	10.5197	14.6114	37.6525	40.6465	44.3141	46.9278	52.620
26	11.1603	15.3791	38.8852	41.9232	45.6417	48.2899	54.052
27	11.8076	16.1513	40.1133	43.1944	46.9630	49.6449	55.476
28	12.4613	16.9279	41.3372	44.4607	48.2782	50.9933	56.892
29	13.1211	17.7083	42.5569	45.7222	49.5879	52.3356	58.302
30	13.7867	18.4926	43.7729	46.9792	50.8922	53.6720	50.703
40	20.7065	26.5093	55.7585	59.3417	63.6907	66.7659	73.402
50	27.9907	34.7642	67.5048	71.4202	76.1539	79.4900	86.661
60	35.5346	43.1879	79.0819	83.2976	88.3794	91.9517	99.607
70	43.2752	51.7393	90.5312	95.0231	100.425	104.215	112.317
80	51.1720	60.3915	101.879	106.629	112.329	116.321	124.839
90	59.1963	69.1260	113.145	118.136	124.116	128.299	137.208
100	67.3276	77.9295	124.342	129.561	135.807	140.169	149.449

Table III
t - Distribution
Values of $t_{\varepsilon, n}$

n	$\varepsilon = 0.4$	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.320	318.310	636.620
2	0.289	0.816	1.866	2.920	4.303	6.965	9.925	14.089	22.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.353	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883

Table III (*continued*)
t - Distribution
Values of $t_{\varepsilon, n}$

n	$\varepsilon = 0.4$	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.951	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.000	3.291

Table IV(A)

***F* - Distribution**

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.05$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50	241.90
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60

Table IV(A) (*continued*)

***F* - Distribution**

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.05$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	3.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Table IV(A) *(continued)*

***F*-Distribution**

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.05$

$n_2 \backslash n_1$	12	15	20	24	30	40	60	120	∞
1	243.90	245.90	248.00	249.10	250.10	251.10	252.20	253.30	254.30
2	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.69	2.60	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13

Table IV(A) (*continued*)

F - Distribution

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.05$

$n_2 \backslash n_1$	12	15	20	24	30	40	60	120	∞
15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	1.01
17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

Table IV(B)

***F* - Distribution**

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.01$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.42	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10,29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94

Table IV(B) (*continued*)

F - Distribution

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0{,}01$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
120	6.85	4.79	3.95	2.48	2.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Table IV(B) (*continue*

***F* - Distribution**

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.01$

$n_2 \backslash n_1$	12	15	20	24	30	40	60	120	∞
1	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00

Table IV(B) (continued)

F - Distribution

Values of $F_{\varepsilon, n_1, n_2}$ where $\varepsilon = 0.01$

$n_2 \backslash n_1$	12	15	20	24	30	40	60	120	∞
15	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

BIBLIOGRAPHY

1. **Baisnab, A. P. and Jas, M ;**
   Elements of Probability and Statistics.
   Tata McGraw-Hill Publishing Co. Ltd., New Delhi.

2. **Bhat, B. R.**
   Modern Probability.
   Wiley Eastern Limited, New Delhi.

3. **Cacoullous, T.**
   Exercises in Probability.
   Narosa Publishing House, New Delhi.

4. **Chatterjee, B. C.**
   General Topology and Elements of Set Theory.
   Dasgupta and Co. Private Limited, Calcutta.

5. **Chung, K. L.**
   A Course in Probability Theory.
   HARCOURT, BRACE and WORLD INC.

6. **Chung, K. L.**
   Elementary Probability Theory with Stochastic Process.
   Narosa Publishing House, New Delhi.

7. **Cramer, H.**
   The Elements of Probability Theory and Some of its Applications.
   John Wiley and Sons. Inc. New York, London, Sydney.

8. **Das, K. P.**
   Sambhabanar Ganitik Tatta O Tahar Proyog.
   Paschimbanga Rajya Pustak Parsad, Calcutta.

9. **Dutta, M. and Pal, S.**
   Introduction to the Mathematical Theory of Probability and Statistics.
   The World Press Private Limited (1963), Calcutta.

10. **Feller, William**
    An Introduction to Probability Theory and its Applications, Vol-I.
    Asia Publishing House, Calcutta.

11. **Gnedenko, B. V.**
The Theory of Probability.
(Translated by George Yankovsky)
Mir Publishers, Moscow.

12. **Goon, A. M. ; Gupta, M. K. and Dasgupta, B.**
An Outline of Statistical Theory (Vol-I).
The World Press Private Limited (1988), Calcutta.

13. **Goon, A. M. ; Gupta, M. K. and Dasgupta, B.**
Fundamentals of Statistics (Vol-I) (Sixth Revised Edition).
The World Press Private Ltd. (1991), Calcutta.

14. **Gupta, Amritava**
Groundwork of Mathematical Probability and Statistics
Academic Publishers, Calcutta.

15. **Gupta, Amritava**
Introduction to Mathematical Analysis.
Academic Publishers, Calcutta.

16. **Hoffman, K. and Kunz, R.**
Linear Algebra.
Prentice Hall.

17. **Kapur, J. N. and Saxena, H. C.**
Mathematical Statistics (Eighth Revised Edition).
S. Chand & Company Ltd, New Delhi.

18. **Kolmogrov, A. N.**
Foundations of the Theory of Probability.
Chelsea Publishing Company, New York.

19. **Leadership Project Committee (Mathematics)**
University of Bombay
Text Book of Algebra.
Tata McGraw-Hill Publishing Company Limited, New Delhi.

20. **Leadership Project Committee (Mathematics)**
University of Bombay
Text Book of Mathematical Analysis.
Tata McGraw-Hill Publishing Company Limited, New Delhi.

21. **Lipschutz, S.**
Set Theory and Related Topics.
Schaums Outlines Series.
McGraw-Hill Book Company, Singapore.

22. **Loeve, Michael**
Probability Theory (2nd Edition).
D. Van Nostrand Company, Inc.

23. **Malik, S. C. and Arora, S.**
Mathematical Analysis.
Wiley Eastern Limited, New Delhi.

24. **Mathi, A. M. and Rathie, P. N.**
Probability and Statistics.
Macmillan India Limited, Delhi.

25. **Medhi, J.**
Statistical Method.
Wiley Eastern Limited, New Delhi.

26. **Mendelson, Elliot**
Boolean Algebra and Switching Circuits.
McGraw-Hill Book Company.

27. **Mood, A. M. ; Greybill, F. A. and Boes, P. C.**
Introduction to the Theory of Statistics.
McGraw-Hill International Editions (Statistics Series)

28. **Mukhopadhyay, Parimal**
Theory of Probability.
New Central Book Agency, Calcutta.

29. **Parthasarathy, K. R.**
Introduction to Probability and Measure.
Macmillan Company of India Ltd. Delhi, 1977.

30. **Rohatgi, V. K.**
An Introduction to Probability Theory and Mathematical Statistics.
Wiley Eastern Limited, New Delhi.

INDEX

(The numbers refer to pages)

ntinue further study with :

MATHEMATICAL STATISTICS

By

Dr. S.K Dey & Dr. S. Sen

ntaining :

- POPULATION AND SAMPLE
- IMPORTANT SAMPLING DISTRIBUTIONS
- ESTIMATION OF PARAMETERS
- COMPUTATION OF SAMPLE CHARACTERISTIC
- INTERVAL ESTIMATION
- BIVARIATE SAMPLES
- TESTING OF HYPOTHESES

—— *the book* ***"Mathematical Statistic"*** *is a sequel to the book* ***"Mathematical Probability"****. This book is primerily designed to serve as a* ***text book*** *for undergraduate students of* ***Mathematics Honours****. It is expected that the book will also be helpful to the students of other disciplines, e.g.* ***Economics, Statistics, Engineering, Commerce and to the students for some competitive examinations****.*

The purpose of the book is to develop the mathematical theory of statistical methods in so far as these are based on the ***concept of probability****.*

Best efforts have been made to treat the contents of the subject in a precise and rigorous manner. At the end of each chapter ***sufficient number of problems have been worked out as illustrative example*** *followed by exercise containing a good number of problems. The problems are selected considering the nature of questions set in University and other examinations.*